Robust Bayesian Regression Model of Centrality and Voltage Stability Index for Power Networks under Nodal Attack

Deepak Kumar Panda[1], Saptarshi Das[1,2], *Member, IEEE*
[1] *Department of Mathematics, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Penryn TR10 9FE, UK.*
[2] *Institute for Data Science and Artificial Intelligence, University of Exeter, Laver Building, North Park Road, Devon, EX4 4QE, UK.*
Email: [1]dp457@exeter.ac.uk, [2]saptarshi.das@ieee.org

***Abstract*—Electrical node centrality for the power networks is an essential parameter to identify the critical nodes under attack. Topological analysis is vital for evaluating the network robustness while electrical characteristics have to be considered to make the analysis consistent for realistic power networks. However, the capacity limit of the power network changes under various nodal attacks. It is essential to find the relationship between the loading margin limit of the power network with the node centrality features, so that appropriate measures can be considered to improve the robustness of the power networks. Thus, voltage stability index (VSI) is defined for every node, and its centrality features are modelled. Robust Bayesian regression is used to model the nodes responsible for a change in loading margin and causing grid blackout. The method has been validated on benchmark complex power networks like reduced Great Britain network, IEEE 57-bus and IEEE 118-bus systems.**

Keywords—Centrality index, nodal attack, robust Bayesian regression, voltage stability index.

I. Introduction

Power grids are critical infrastructure in modern society. Extreme events or component failure can cause a widespread blackout or a massive service interruption in a large area causing severe socio-economic consequences [1]. In order to prevent power blackouts, it is essential to identify critical nodes and branches in the power network where network centrality measures with electrical and topological features are mainly used. Voltage instability of the power grid can cause a major blackout. As defined in [2], voltage stability is the ability of the power system to maintain its voltage such that, with the increase in load, both power and voltage are controllable. It is the ability to maintain steady-state voltage after subjected to a disturbance. As given in [3], in order to avoid voltage instability, weak buses or lines are needed to be improved by distributed generations or voltage supporting equipments. The Voltage stability index (VSI) helps in finding the weak nodes/links in the power networks [4], [5].

A. Previous works in Power Network Centrality Measures

The robustness analysis of the power network to topological features is conducted in [6] using characteristic path length, node connectivity loss and blackout size. Similarly, in [7], power grid evolution based on the cost of the distribution grid is evaluated. In [8], the network reconfiguration efficiency is proposed based on the network node degree distribution and the clustering coefficient of the network. However, the analysis, which is based on purely topological features, can be misleading, as it does not necessarily capture the complexity of the power grid. The node centrality concept, based on AC power flow model [9] and betweenness centrality of the graph which is based on power network line admittance [10], have been utilized here to study the blackout model for the power grids. Several indices of the power networks are introduced in [11] to identify the critical nodes in the skeleton network configuration which captures the topological and electrical characteristics of the power grid. Multi-criteria decision making (MCDM) strategies are leveraged using these features to identify the critical nodes in the network along with studying the correlation structures between different complex network features of the power grid.

B. Previous Works on Voltage Stability Index

Voltage stability of power grids is affected by the static and dynamic characteristics of the loads [12]. Continuation power flow (CPF) method [13] provides an advantage over the traditional power flow technique in terms of convergence which traces the solution of the loading margin path with the change in the bus voltage. It is a widely accepted measure as it provides the distance of the power system from voltage collapse. The VSI is proposed in [14], [15] using the local voltage phasors, as measured by the phasor measurement units (PMUs). In [15], VSI based on Tellegen's theorem is proposed, which is simple, computationally tractable and easy to implement in a wide-area monitoring environment and the control centers. Various methods for VSI is mentioned in [16] based on the power network buses and branches. The VSI is derived in the context of voltage phasors, nodal current, equivalent change in load bus impedance and the Thevenin's impedance of the network. However, the concept of VSI based on nodal attack has not been defined in the previous literature, which is the main focus of this paper. It is essential to study the change in the properties of the power grid when a node is disconnected from the network due to an attack as shown in [17]–[19]. Hence, the concept of VSI is defined in this paper, which is the relative change in the loading margin of the power network before voltage collapse occurs.

C. Contributions of This Paper

In this paper, the concept of VSI has been introduced in terms of the nodal attacks on the grid. The measure is obtained for all the nodes, which are attacked in the power grid, causing a change in network loading margin. The CPF has been used to find the given measures. The VSI, based on the local voltage measurements, as discussed in the previous section, are not reliable enough to capture the power grid

978-1-7281-8398-5/21/$31.00 ©2021 IEEE

resilience in the case of a nodal attack. In addition to this, VSI for the nodes is modelled along with the indices defined in [11]. This study is conducted using a robust Bayesian regression model, where a customized likelihood function is defined for the nodes causing voltage collapse under nodal attack. There are very few literatures available, which conduct a topological analysis of voltage collapse. In [20], some theoretical insight has been provided on the grid structure, which influences voltage collapse. However, the analysis is conducted on a simplified power flow model based on only active power demands. Hence, the analysis fails near the voltage collapse points. Hence, robust regression methods can improve the analysis to find the dependencies of the VSI with the topological and electrical features of the grid. The model has been validated on three-benchmark complex networks viz. IEEE 57-bus, IEEE 118-bus systems and the reduced Great Britain power network. Hence the main contributions of this paper can be summarized as follows:

- Defining the voltage stability index of the power grid, based on the relative change in the loading margin due to nodal attacks.
- Fitting a robust Bayesian regression model to find the dependencies of the network feature with VSI using of Hamiltonian Monte Carlo (HMC) sampling technique for the customized likelihood function defined for each node causing a blackout in the power network.

II. Nodal Centrality Measures

The *N*-node power network can be represented as an $N\times N$ adjacent matrix $\{A_{ij}\}$ where the elements of A_{ij} represents the connectivity between different nodes. Hence, the node centrality measures can be defined based on the adjacency matrix.

A. Electrical Degree Centrality

In the power network, the nodes having higher generation capacities and load demands are very critical. It represents a high amount of power going in and coming out of a node in the power network. Electrical degree centrality (EDC) for a power network having nodes *N* captures this aspect for node *i* as follows [21]:

$$I_{EDC_i}=\left(\sum_{j\in\Gamma(i)}A_{ij}\Big/(N-1)\right)e^{-\rho}. \tag{1}$$

Here, $\rho=\left[\left(\left(s_{G,i}+s_{L,i}\right)/s_{G,\max}\right)-1\right]$, where $\{s_{G,i},s_{L,i}\}$ represent the generation and load capacity of node *i,* while $s_{G,\max}$ represents the maximum generation capacity amongst the generators. $\Gamma(i)$ represent the adjacent nodes of node *i*.

B. Electrical Closeness Centrality

It is essential during a blackout that the restoration path of a power network contains a smaller number of transmission lines and transformer branches, as it improves the restoration time with less charging capacitance in the transmission lines and reduces the possibility of over-voltage. The electrical closeness centrality (ECC) highlights this aspect of the power network [21], which is defined as:

$$I_{ECC}(i)=(N-1)\Big/\sum_{j=1,j\neq i}^{N}d_{\min,ij}^{Q_C}, \tag{2}$$

where, $d_{\min,ij}^{Q_C}$ represents the minimum total capacitance of the transmission line between node *i* and *j*.

C. Electrical Betweenness Centrality

In the power flow analysis, the power flow might not always happen with the shortest route between the generator and the load. The electrical betweenness centrality (EBC) index is introduced in [22], which can be defined as:

$$I_{EBC}(i)=\sum_{g\in\Gamma_{SE}}\sum_{l\in\Gamma_{SK}}\left(\sum_{j\in\Gamma(i)}\left|f_{ij}^{gl}\right|\right), \tag{3}$$

where, $f_{ij}^{gl}=\left[\left(Z_{ig}^{eq}-Z_{il}^{eq}\right)-\left(Z_{jg}^{eq}-Z_{jl}^{eq}\right)\right]\Big/x_{ij}$ and $Z_{ig}^{eq}=\left(z_{ii}-z_{ig}\right)-\left(z_{ig}-z_{gg}\right)$.

Here, Γ_{SE} and Γ_{SK} represent power and load set respectively. Also, f_{ij}^{gl} represents the power transferred between source node (*g*) and load node (*l*) with the transmission line connecting node *i* and *j*. x_{ij} represents the reactance of the transmission line L_{ij}. The parameters Z_{ig}^{eq}, z_{ig}, z_{ii} and z_{gg} in (3) represents equivalent impedance between node *i* and source node *g*, transfer impedance between *i* and source node *g*, and driving point impedances of node *i* and *g* respectively. This is calculated using the power transfer distribution factor, as defined in [22]. Higher value of $I_{EBC}(i)$ signifies higher role of node *i* in the power flow between the nodes in the network and vice-versa.

D. Eigenvector Centrality

Let us consider χ and $e=[e_1,e_2,\ldots,e_N]^T$ being the dominant eigenvalue and eigenvector for the $N\times N$ adjacency matrix $\{A_{ij}\}$ with *N* nodes. Thus following [21], we get:

$$\chi e_i=\sum_{j=1}^{N}A_{ij}e_j,\quad i=1,2,\ldots,N. \tag{4}$$

Hence, the eigenvector centrality (EVC) of the node can be defined as:

$$I_{EVC}(i)=\frac{1}{\chi}\sum_{j=1}^{N}A_{ij}e_j. \tag{5}$$

This index is essential for the importance of a node in the topological perspective.

E. Network Efficiency Centrality

The network efficiency considering the electrical characteristics is given as:

$$E_B=\frac{1}{N(N-1)}\sum_{k,j\in V_N}\frac{1}{d_{\min,kj}^{X_L}}, \tag{6}$$

where, $d_{\min,kj}^{X_L}$ represents the shortest path with the minimum number of transmission lines and transformer branches to line reactance between node *k* and *j*. The V_N represents the set of nodes in the power network. Thus, network efficiency centrality (NEC) for node *i* is defined as the relative change in the efficiency after the respective node is removed, which is defined as:

$$I_{NEC}(i)=\left(E_B-E_{B,i}\right)/E_B, \tag{7}$$

where,

$$E_{B,i} = \frac{1}{(N-1)(N-2)} \sum_{\substack{k,j \in V_N \\ k \neq i,\, j \neq i}} \left(1/d_{\min,kj}^{X_L}\right). \quad (8)$$

Here in (8), $E_{B,i}$ represents the network efficiency after node i is removed.

F. Rate of Change of Spanning Trees

After removing specific key nodes from the graph, the number of spanning trees gets reduced, hence making the network disconnected [21]. The change in spanning-tree (CST) centrality can be defined as the relative change in the number of spanning trees when node i is removed which is represented as:

$$I_{CST}(i) = 1 - \left[\tau(G-i)/\tau(G)\right], \quad (9)$$

where, $\tau(G)$ represents the number of spanning trees in the network or graph G while $\tau(G-i)$ represents the number of spanning trees with the node i removed. Higher the value of $I_{CST}(i)$, the graph becomes more disconnected with the removal of i and vice-versa.

G. Rate of Change of Network Closeness Centrality

The importance of a node in the power network can be calculated as the relative change in the network closeness centrality (NCC) after node i is removed [21], as it highlights the impact of restoring a node in the power network by its contraction method [8]. It can be defined as:

$$I_{NCC}(i) = (N-1) \Bigg/ \left(2 \sum_{1 \le k < j \le N} d_{\min,kj}^{X_L} \right), \quad (10)$$

N represents the number of nodes in the network. Higher value of $I_{NCC}(i)$ means that node i, is closer to other nodes in the network and vice-versa.

III. Voltage Stabilty Index

As shown in Figure 1, the first task is to find the loading margin of the power network, which is found by solving the continuation power flow or CPF [13], which is reformulated power flow technique solved using continuation techniques. The techniques involve a predictor-corrector method to incorporate a load parameter so that divergence and the error due to singularity of the power flow Jacobian are not encountered. When it comes to CPF, the existing power flow model is augmented with a constant power model. Let us consider be the load parameter such that,

$$0 \le \lambda \le \lambda_{\text{crit}}, \quad (11)$$

where, λ_{crit} corresponds to critical load that can cause voltage collapse. Here, $\lambda = 0$ represents the base load condition. Hence, the active and reactive power flow equations are:

$$P_{Gi} - P_{Li} - P_{Ti} = 0, Q_{Gi} - Q_{Li} - Q_{Ti} = 0. \quad (12)$$

where, $P_{Ti} = \sum_{j=1}^{N} V_i V_j y_{ij} \cos\left(\delta_i - \delta_j - \nu_{ij}\right)$, is the injected real power and $Q_{Ti} = \sum_{j=1}^{N} V_i V_j y_{ij} \sin\left(\delta_i - \delta_j - \nu_{ij}\right)$ is the injected reactive power in the i^{th} bus. $\{V_i \angle \delta_i, V_j \angle \delta_j\}$ represents the voltage at i^{th} and j^{th} bus respectively and $y_{ij} \angle \nu_{ij}$ represents the admittance between i^{th} and j^{th} bus for a N bus power network. The continuation solutions are obtained by varying the load $\{P_{Li}, Q_{Li}\}$ and generator $\{P_{Gi}, Q_{Gi}\}$ at each bus with the load parameter λ. They are parameterized as:

$$P_{Li} = P_{Li_0} + \lambda\left(k_{Li} S_{\Delta base} \cos\psi_i\right), Q_{Li} = Q_{Li_0} + \lambda\left(k_{Li} S_{\Delta base} \sin\psi_i\right),$$
$$P_{Gi} = P_{Gi_0} + \lambda\left(k_{Gi} S_{\Delta base} \cos\psi_i\right), Q_{Gi} = Q_{Gi_0} + \lambda\left(k_{Gi} S_{\Delta base} \sin\psi_i\right). \quad (13)$$

where, $\{P_{Li_0}, Q_{Li_0}\}$ and $\{P_{Gi_0}, Q_{Gi_0}\}$ represent the base active and reactive power for load and generator respectively. Parameters $\{k_{Li}, k_{Gi}\}$ represent the respective load and generator change at i^{th} bus, with modifications in λ. Here, $S_{\Delta base}$ is the base apparent power for the power network and ψ_i represents the load angle at i^{th} bus. Substituting (13) in (12), yields the following set of equations:

$$P_{Gi_0}\left(1 + \lambda k_{G_i}\right) - P_{Li_0} - \lambda\left(k_{L_i} S_{\Delta base} \cos\psi_i\right) - P_{T_i} = 0,$$
$$Q_{Gi_0} - Q_{Li_0} - \lambda\left(k_{L_i} S_{\Delta base} \sin\psi_i\right) - Q_{T_i} = 0. \quad (14)$$

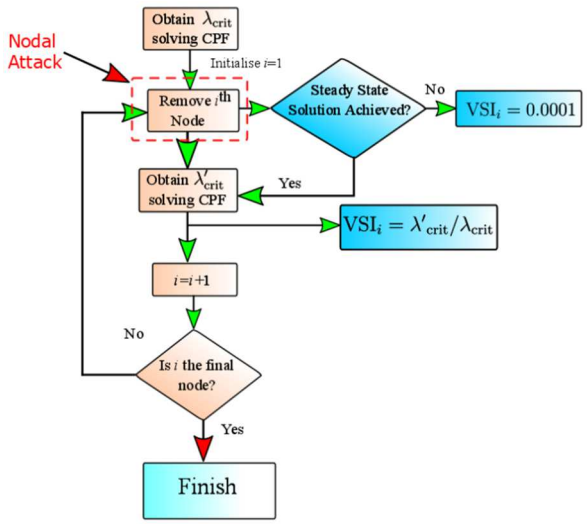

Figure 1: Schematic to calculate VSI for all the nodes in the power network.

The solution set $\{V, \delta, \lambda\}$ is obtained after solving (14). The parameterization of $\{V, \delta, \lambda\}$ in (11) helps in quantifying the solution obtained between successive iterations using the continuation step-size parameter σ. The pseudo-arc length parameterization is used in [23], where the solution $\{V^{j+1}, \delta^{j+1}, \lambda^{j+1}\}$ is constrained to lie on the hyperplane of the tangent of $\{V^j, \delta^j, \lambda^j\}$ which is represented as:

$$p^j(x,\lambda) = \left(\begin{bmatrix} V \\ \delta \\ \lambda \end{bmatrix} - \begin{bmatrix} V^j \\ \delta^j \\ \lambda^j \end{bmatrix} \right)^T \bar{z}^j - \sigma^j = 0, \quad (15)$$

Here, $\bar{z}^j$ represents the normalized tangent vector of the solution $\{V^j, \delta^j, \lambda^j\}$ and σ^j is the adaptive step size parameter. The λ_{crit} is obtained when the following voltage collapse condition i.e.

$$d\lambda\big|_{\lambda=\lambda_{\text{crit}}} = 0, \tag{16}$$

is satisfied with the iteration given in (15). When i^{th} node is removed as shown in Figure 1, the topology of the network changes with the change in the admittance matrix constituting the power flow as shown in the following equations:

$$\begin{aligned} &P_{Gi_0}\left(1+\lambda k_{G_i}\right) - P_{Li_0} - \lambda\left(k_{L_i} S_{\Delta base}\cos\psi_i'\right) \\ &\quad - \sum_{j=1}^{N-1} V_i V_j y_{ij}' \cos\left(\delta_i - \delta_j - \nu_{ij}'\right) = 0, \\ &Q_{Gi_0} - Q_{Li_0} - \lambda\left(k_{L_i} S_{\Delta base}\sin\psi_i'\right) \\ &\quad - \sum_{j=1}^{N-1} V_i V_j y_{ij}' \sin\left(\delta_i - \delta_j - \nu_{ij}'\right) = 0. \end{aligned} \tag{17}$$

Hence, the solution set $\{V', \delta', \lambda'\}$ is obtained by solving (17) considering the parameterization defined in (15) till λ_{crit}' is satisfied. However, there will be situations while solving (17) for some σ, where equation (16) will not be satisfied, hence the steady state solution of (17) is not obtained. Hence, for those nodes a low value of VSI_i is obtained, as no steady state solution for the network is obtained for the network when those nodes are removed from the network. We now define the voltage stability index of that particular node as:

$$\text{VSI}_i = \begin{cases} 0.0001 & \text{if equation (16) is satisfied,} \\ \lambda_{\text{crit}}'/\lambda_{\text{crit}} & \text{if equation (16) is not satisfied.} \end{cases} \tag{18}$$

If $\text{VSI}_i > 1$ then, the loading margin limit of the network increases. Similarly, if $\text{VSI}_i < 1$, the network can handle smaller perturbations and the network stability decreases. A smaller value of VSI_i indicates that *i* is a critical node of the network and removing it can cause a network-wide blackout. The statistical method to model the VSI in (18) with the graph indices defined in section 2 has been explained in the next section. The parameters $\{\lambda_{\text{crit}}, \lambda_{\text{crit}}'\}$ are obtained by running the CPF routine in MATPOWER [24] for three different benchmarks power networks. The data for the reduced GB network is obtained from [25]. The nodes signifying the increase and decrease in loading margin of the network, along with the ones leading to complete blackout are shown in Figure 2. It is evident that the nodes that are topologically adjacent to each other have similar VSI characteristics.

IV. Robust Bayesian Regression with Custom Likelihood Function

It is crucial to find out the relationship between the node centrality measures defined in Section II with the VSI defined in the previous section. As we see in Figure 2, for some of the nodes the VSI_i is low, leading to a complete blackout. It is essential to model these nodes from their respective node centrality measures. These values are modelled as an outlier, and the robust regression model is fitted. It is done by placing a *t*-distribution prior to the observed data [26] and sampling the posterior accordingly. However, as we see in Figure 2, it is difficult for the *t*-distributed priors to capture the nodes which have similar outlier values. Hence it is vital for customizing the likelihood to label the network nodes causing a complete network blackout when attacked [27]. Those observed indices are modelled with a binary indicator, to segregate them from the normal nodes where we get a definitive VSI. The outlier binary parameter is modelled with larger deviations from the usual Gaussian noise estimates which are generally known as the "sigma clipping" [28].

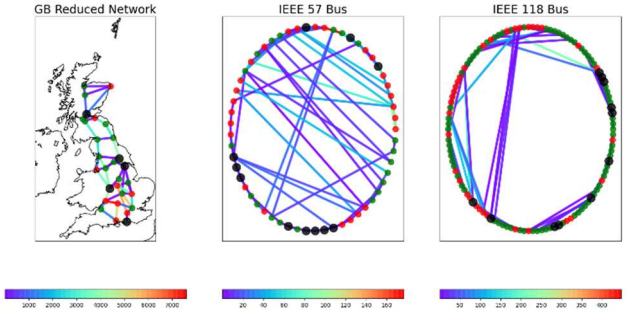

Figure 2: Voltage stability index obtained for complex power networks for GB-reduced network (29 nodes), IEEE 57-bus and IEEE 118-bus systems. Meaning of the colors are: Red represent VSI_i<1 (decrease in loading margin), Green represent VSI_i>1 (increase in loading margin), Black represent $\text{VSI}_i = 1$ (steady state limit not found). The edges represent active power flow between the nodes with the colourmap signifying its magnitude.

Let us consider the VSI with the centrality measures as

$$\text{VSI} = \varkappa\boldsymbol{\beta}, \tag{19}$$

where, $\varkappa = [I_{EDC} \quad I_{ECC} \quad I_{EBC} \quad I_{EVC} \quad I_{NEC} \quad I_{NCC} \quad I_{CST}]$ and $\boldsymbol{\beta} = [\beta_1 \quad \beta_2 \quad \dots \quad \beta_7]^T$. There are N nodes with the binary integers q_i for each nodes, where it is zero if the node causes a blackout and one if (11) converges. We set prior probability $\mathcal{P}_b$ for each node. Hence, we obtained $(N+3)$ number of extra parameters which are later marginalized to obtain $\boldsymbol{\beta}$ at the end with the observed I data. Hence, we write the likelihood function as:

$$\begin{aligned} \text{L} &= p\left(\{\text{VSI}_i\}_{i=1}^N \middle| \boldsymbol{\beta}, \{q_i\}_{i=1}^N, \mathcal{Y}_b, \mathcal{V}_b, I\right) \\ &= \prod_{i=1}^N \left[p_{\text{fg}}\left(\{\text{VSI}_i\}_{i=1}^N \middle| \boldsymbol{\beta}, I\right)\right]^{q_i} \left[p_{\text{bg}}\left(\{\text{VSI}_i\}_{i=1}^N \middle| \mathcal{Y}_b, \mathcal{V}_b, I\right)\right]^{[1-q_i]}, \end{aligned} \tag{20}$$

where, $p_{\text{fg}}(.)$ and $p_{\text{bg}}(.)$ represent the distribution from where the VSI for the non-blackout and blackout nodes are sampled. In (20), $\{\mathcal{Y}_b, \mathcal{V}_b\}$ represent the mean and variance of the VSI of critical nodes. In order to separate the different nodes in $\{q_i\}_{i=1}^N$, the binomial probability distribution function (pdf) $\mathcal{P}_b$ is used as:

$$\begin{aligned} &p\left(\boldsymbol{\beta}, \{q_i\}_{i=1}^N, \mathcal{P}_b, \mathcal{Y}_b, \mathcal{V}_b \middle| I\right) = \\ &\qquad p\left(\{q_i\}_{i=1}^N \middle| \mathcal{P}_b, I\right) \cdot p\left(\boldsymbol{\beta}, \mathcal{P}_b, \mathcal{Y}_b, \mathcal{V}_b \middle| I\right), \\ &p\left(\{q_i\}_{i=1}^N \middle| \mathcal{P}_b, I\right) = \prod_{i=1}^N \left[1-\mathcal{P}_b\right]^{q_i} \mathcal{P}_b^{[1-q_i]}. \end{aligned} \tag{21}$$

Applying this principle in (20) yields:

$$\begin{aligned} \text{L} = \prod_{i=1}^N &\left[\frac{1-\mathcal{P}_b}{\sqrt{2\pi\sigma_{VSI_i}^2}} \exp\left(-\frac{(VSI_i - \beta x_i)}{\sigma_{VSI_i}^2} \right) \right] \\ &+ \left[\frac{\mathcal{P}_b}{\sqrt{2\pi\left[\mathcal{V}_b + \sigma_{VSI_i}^2\right]}} \exp\left(-\frac{\left[VSI_i - \mathcal{Y}_b\right]^2}{2\left[\mathcal{V}_b + \sigma_{VSI_i}^2\right]} \right) \right]. \end{aligned} \tag{22}$$

In order to find the posterior probability distribution of the coefficients $\boldsymbol{\beta}$ in (19) using the nodal features and the VSI, marginalization is carried out accounting for the covariance and other parameters $\{P_b, \gamma_b, \nu_b\}$ and given as:

$$p\left(\boldsymbol{\beta}\middle|\{\mathrm{VSI}_i\}_{i=1}^{N}\right) = \int \mathrm{d}\{q_i\}_{i=1}^{N}\, \mathrm{d}P_b\, \mathrm{d}\gamma_b\, \mathrm{d}\nu_b\, p\left(\boldsymbol{\beta}, \{q_i\}_{i=1}^{N}, \gamma_b, \nu_b, I\right). \tag{23}$$

Markov Chain Monte Carlo (MCMC) sampling for numerical integration is to be used next to find the high dimensional marginal distributions in (23). We here use the Hamiltonian Monte Carlo sampler [29], which is useful in high dimensional Bayesian inference. The numerical results have been explained in the next section.

V. Results and Discussions

The complex network measures discussed in Section II are obtained from the brain connectivity toolbox [30], and the graph and network functions in MATLAB. The nodal measures are plotted in Python package Seaborn [31] along with several categories of VSI for different power grids, which are shown in Figure 3. Here, we observe a distinct pattern with the change of the number of spanning trees for the three VSI cases, considering different power grid topologies. Distinctness is also observed in Figure 3 for the case of EBC and NCC since the distributions do not overlap for the three power grids considered here, which confirms that the fractions of the nodes involved in power flow are different for the three different benchmark power grids.

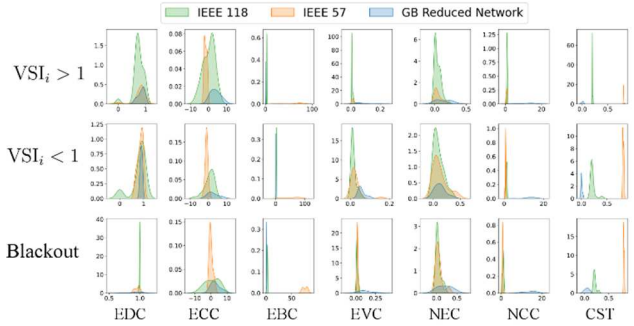

Figure 3: Univariate distribution of the node centrality measures for the three benchmark power networks: reduced GB network, IEEE 57-bus and IEEE 118-bus system with different VSI categories. The distributions represent complex network local features associated with each node. (Glossary of abbreviations: EDC- Electrical Degree Centrality, ECC- Electrical Closeness Centrality, EBC- Electrical Betweenness Centrality, EVC-Eigen Vector Centrality, NEC-Network Efficiency Centrality, NCC-Network Closeness Centrality, CST-Change in Spanning Trees).

The closeness factors of the nodes are also quite distinct for the power grids considered here. We observe a distinct overlap in the three parameters - EDC, ECC and NEC. However, it is also observed that the nodes, which causes blackout, have low EDC, as evident from the network plot shown in Figure 2. It is interesting to check the relation of the network centrality indices with the VSI, which is obtained by running a robust Bayesian regression model with the likelihood defined in (22) implemented using the PyMC3 package in Python [32]. Since we are using binary indicators $\{q_i\}_{i=1}^{N}$ along with the continuous variables for modelling the characteristics of the nodes, it is helpful we scale the continuous variables by two standard deviations for efficient regression modelling [33]. The prior distribution for $\boldsymbol{\beta}$ is considered to be normal as $\mathrm{N}(0,10)$. As given in [34], the prior for the standard deviation of VSI $\sigma_{\hat{\mathrm{VSI}}}$ is considered to be inverse gamma distribution with parameters $\bar{\alpha}=3$ and $\bar{\beta}=0.5$. The prior for $\sigma_{\hat{\mathrm{VSI}}}^{\mathrm{out}}$ is considered to be half-normal distribution as discussed in [27]. The standard deviations $\sigma_{\hat{\mathrm{VSI}}}$ and $\sigma_{\hat{\mathrm{VSI}}}^{\mathrm{out}}$ are used along with the $\boldsymbol{\beta}$ to model the VSI of the node, causing network blackout. The binary index of the distribution P_b is obtained from the Bernoulli distribution, which takes a uniform prior between 0 and 0.5 suggesting the fraction of the nodes causing the entire power grid blackout when it is attacked. The No-U turn sampler (NUTS) [29] which adaptively sets the path length of the Hamiltonian Monte Carlo sampler, is utilized here for finding the posterior distributions of the regression coefficients (19).

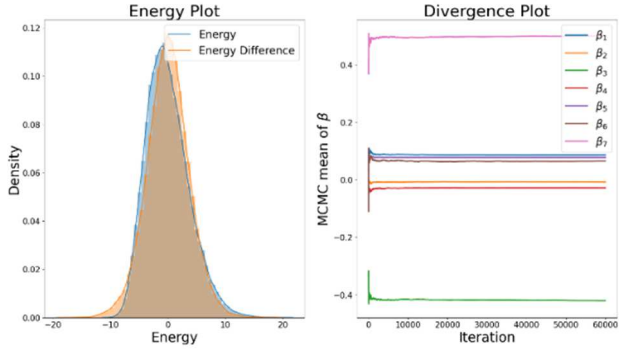

Figure 4: HMC sampling diagnostics using energy and divergence plot.

The sampling algorithm for the robust Bayesian regression modeling has been run on a 64-bit Windows PC with Intel Core i5-8500 CPU, 3 GHz processor with six parallel cores. The `sample()` function is utilized from the PyMC3 package in the sampling process, by running multiple parallel chains on the 6 parallel CPU cores to speed up the computationally expensive sampling process. Four different chains are used for the inference with the symplectic integrator to generate the trajectories, thus preserving the phase space volume of Hamiltonian transition. Random walk Metropolis-Hastings (MH) is utilized for adaptive step length of the symplectic integrator, thus reducing the bias in the resultant Hamiltonian transition and yielding the samples from target distribution [29]. The target acceptance probability of 0.95 for the Metropolis-Hastings (MH) sampler with symplectic integrator is used in the sampling process. We have used 15000 samples for each chain, and it was assigned to each of the 6 parallel cores of the CPU. The time for parallel computing has also been compared with respect to the sampler running on a single core. The computing time was ~1159 seconds = 19.3 mins for the chains to run on a single core while it took ~380 seconds = 6.3 mins when all the 6 cores of the CPU were utilized, thus indicating a multicore vs. single core speedup factor of 3.05X. The diagnostics of the sampling are shown in Figure 4. As observed from Figure 5, the distribution of the VSI is similar for all the buses, hence a resultant robust statistical model is formulated based on the likelihood in (22) for the centrality features and voltage stability index of the nodes.

The trajectory of the data from the HMC sampler is confined in an energy set. Hence, the chains decouple into a deterministic and stochastic exploration of the samples amongst level sets. As we observe from Figure 4 first subplot, that the energy difference between the level sets is similar to the samples between the level sets suggesting that the random walk of the chain has explored the marginal energy distributions efficiently. In the second subplot of Figure 4, we observe that mean of the parameter is converging to a steady state value. It suggests that the transition energy between the samples remains finite with sufficient exploration with no divergence. This is due to the high acceptance rate specified in the built-in MH sampler which explores the parameter space well but makes the convergence slower.

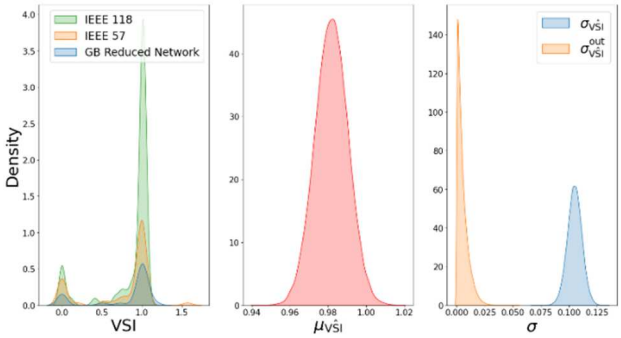

Figure 5: Distribution of the VSI data in (18) for the benchmarks power networks with mean and standard deviation obtained using HMC sampling.

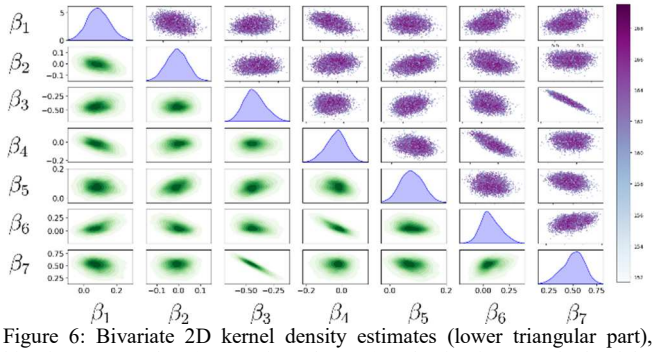

Figure 6: Bivariate 2D kernel density estimates (lower triangular part), bivariate scatter plots (upper triangular part) and univariate marginals of the posterior distributions of the regression coefficients in (19) amongst the nodal parameters and VSI, obtained using the HMC sampler. The scatterplot is scaled according to the cumulative loglikelihood function given in (22).

Hence, it makes the sampling process slower but negates any divergence occurring due to infinite energy occurred during the transitions. The Gelman-Rubin statistic [35] for $\boldsymbol{\beta}$ is found to be 1.01 which confirms the efficiency of the sampling process used. The samples from the posterior are obtained from the converged chains and are shown in Figure 5. We observe that the $\mu_{\hat{VSI}}$ captures the peak obtained from the original VSI data for the different buses. Using the Bayesian regression model in (19), the variance of the estimated coefficients $\boldsymbol{\beta}$ can be mapped on to the observed variance of VSI for those nodes, where the steady state loading margin (λ'_{crit}) of the power network is not obtained on nodal attack. The VSI of the remaining nodes causing blackout is effectively modelled by $\sigma^{out}_{\hat{VSI}}$. However, the posterior spread in $\sigma^{out}_{\hat{VSI}}$ is thinner as compared to the $\sigma_{\hat{VSI}}$ which effectively models the low VSI of the nodes causing where steady state loading margin (λ'_{crit}) of the network is not obtained on nodal attack. The $\mu_{\hat{VSI}}$ in Figure 5 is similar to the VSI obtained from the grid as given in the univariate distribution in Figure 3. Now we examine the dependence of the centrality measures with the VSI by analyzing the marginal univariate and pairwise bivariate posterior distributions of $\boldsymbol{\beta}$ as shown in Figure 6.

We observe from Figure 4, that the mean of β_7 has the highest positive value while that of β_3 has the highest negative value suggesting that the VSI measure of the nodes increase with an increase in the spanning trees. However, an increase in the EBC reduces the VSI of the nodes. It physically means that if the nodes, which generally forms a part of the shortest power flow route, is removed, then it increases the load margin of the power network. However, loading margin of the network reduces if the nodal attack makes the graph disconnected. Hence, we can conclude that the VSI is related to the closeness of the nodes in the graph. We observed from Figure 6 that the relationship between coefficients β_3 and β_7 are highly negatively correlated which signifies the complementary nature of the EBC and CST on the VSI. This indicates that on removing the node, which plays an essential role in the power flow between the other nodes in the network, can make the network less disconnected. This also signifies that the power flow is not dependent on the removal of certain nodes in the network. We also observe a similar moderate negative correlation between coefficients β_4 and β_6. An increase in the network closeness increases the VSI of the nodes, while the nodes that are on the higher end of the eigenvalue spectrum reduces the VSI. The relationship of the network closeness with VSI is also evident with the network plots in Figure 2, where the nodes having high VSI are close to each other topologically. This is observed in the Great Britain map in Figure 2 where areas in Scotland like Glasgow and Inverness and areas in Northern England like Newcastle, Lancaster, Manchester and Leeds fall under the similar voltage stability category. Places in Midlands and South England like Cambridge, Oxford, Gloucester and Southampton have similar voltage stability regions where on nodal attack, the VSI decreases. Larger positive range of values of β_7 in Figure 6 suggests that $I_{NCC}(i)$ has higher influence on VSI as compared to other 6 nodal parameters. It also means that when node *i,* which is electrically closer to other nodes in the network, is attacked, it causes the increase in steady state loadability margin. This observation is also supported in Figure 2, where the nodes with higher VSI are electrically closer to each other. We also observe a negative range of values for β_3 in Figure 6 suggesting that $I_{EBC}(i)$ has negative influence on VSI. This suggests that when node *i*, which plays an essential role in power flow in the network, is attacked, can cause a decrease in the steady state loadability margin of the network.

In the regression model used here, only nodal properties are considered, future works will also incorporate the topological properties of the network like small worldness, modularity, transitivity etc. to check the dependency of VSI with other parameters. The calculation of VSI for larger networks is expensive. Hence, a sampling strategy for nodal

attack can be considered to select the certain nodes to find the appropriate regression model. The effect of scalability of such analysis on larger complex networks is also another potential topic of interest. Further analysis may also consider including other centrality measures, which contribute to the grid failure along with cascaded attacks.

VI. Conclusions

We use a robust Bayesian regression model to find the relationship between the nodal centrality measures and VSI. VSI is assigned based on the relative loading margin of the network under attack. A Bayesian regression model with customized likelihood function has been used to find the posterior distributions of the relationships between the nodal centrality measures and the VSI. The customized likelihood involves separate modelling of the nodes with low VSI. The posterior distributions indicate that the nodes which are in the path of shortest power flow within the network along with the nodes which are in the higher end of the eigenvalue spectrum play a significant role in deteriorating the VSI of the network. However, the loading margin of the network improves if the network closeness and centrality decrease after nodal attack. This suggests that the community structure of the network is robust to nodal attacks.

References

[1] Z. Lin, F. Wen, and Y. Xue, "A restorative self-healing algorithm for transmission systems based on complex network theory," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 2154–2162, 2016.

[2] P. Kundur, J. Paserba, V. Ajjarapu, G. Andersson, A. Bose, C. Canizares, N. Hatziargyriou, D. Hill, A. Stankovic *et al.*, "Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1387–1401, 2004.

[3] R. M. M. Pereira, C. M. M. Ferreira, and F. M. Barbosa, "Comparative study of STATCOM and SVC performance on dynamic voltage collapse of an electric power system with wind generation," *IEEE Latin America Transactions*, vol. 12, no. 2, pp. 138–145, 2014.

[4] Y. Amrane, M. Boudour, and M. Belazzoug, "A new optimal reactive power planning based on differential search algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 551–561, 2015.

[5] I. Musirin and T. A. Rahman, "Estimating maximum loadability for weak bus identification using FVSI," *IEEE Power Engineering Review*, vol. 22, no. 11, pp. 50–52, 2002.

[6] P. Hines, E. Cotilla-Sanchez, and S. Blumsack, "Do topological models provide good information about electricity infrastructure vulnerability?," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 20, no. 3, p. 033122, 2010.

[7] G. A. Pagani and M. Aiello, "Power grid complex network evolutions for the smart grid," *Physica A: Statistical Mechanics and its Applications*, vol. 396, pp. 248–266, 2014.

[8] Y. Liu and X. Gu, "Skeleton-network reconfiguration based on topological characteristics of scale-free networks and discrete particle swarm optimization," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 1267–1274, 2007.

[9] B. Liu, Z. Li, X. Chen, Y. Huang, and X. Liu, "Recognition and vulnerability analysis of key nodes in power grid based on complex network centrality," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 3, pp. 346–350, 2017.

[10] M. Mureddu, G. Caldarelli, A. Damiano, A. Scala *et al.*, "Islanding the power grid on the transmission level: less connections for more security," *Scientific Reports*, vol. 6, p. 34797, 2016.

[11] Z. Lin, F. Wen, H. Wang, G. Lin, T. Mo, and X. Ye, "CRITIC-based node importance evaluation in skeleton-network reconfiguration of power grids," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 2, pp. 206–210, 2017.

[12] M. K. Pal, "Voltage stability conditions considering load characteristics," *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 243–249, 1992.

[13] V. Ajjarapu and C. Christy, "The continuation power flow: a tool for steady state voltage stability analysis," *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 416–423, 1992.

[14] Y. Wang, W. Li, and J. Lu, "A new node voltage stability index based on local voltage phasors," *Electric Power Systems Research*, vol. 79, no. 1, pp. 265–271, 2009.

[15] I. Smon, G. Verbic, and F. Gubina, "Local voltage-stability index using Tellegen's theorem," *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1267–1275, 2006.

[16] J. Modarresi, E. Gholipour, and A. Khodabakhshian, "A comprehensive review of the voltage stability indices," *Renewable and Sustainable Energy Reviews*, vol. 63, pp. 1–12, 2016.

[17] M. A. S. Monfared, M. Jalili, and Z. Alipour, "Topology and vulnerability of the Iranian power grid," *Physica A: Statistical Mechanics and its Applications*, vol. 406, pp. 24–33, 2014.

[18] P.-Y. Chen and A. O. Hero, "Assessing and safeguarding network resilience to nodal attacks," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 138–143, 2014.

[19] H. Cetinay, K. Devriendt, and P. Van Mieghem, "Nodal vulnerability to targeted attacks in power grids," *Applied Network Science*, vol. 3, no. 1, p. 34, 2018.

[20] J. W. Simpson-Porco, F. Dörfler, and F. Bullo, "Voltage collapse in complex power grids," *Nature Communications*, vol. 7, no. 1, pp. 1–8, 2016.

[21] S. Fortunato, G. Mangioni, R. Menezes, and V. Nicosia, *Complex Networks*. New York, NY, USA: Springer, 2009.

[22] F. Wenli, H. Ping, and L. Zhigang, "Multi-attribute node importance evaluation method based on Gini-coefficient in complex power grids," *IET Generation, Transmission & Distribution*, vol. 10, no. 9, pp. 2027–2034, 2016.

[23] H. Mori and S. Yamada, "Continuation power flow with the nonlinear predictor of the Lagrange's polynomial interpolation formula," in *IEEE/PES Transmission and Distribution Conference and Exhibition*, 2002, vol. 2, pp. 1133–1138.

[24] R. D. Zimmerman, C. E. Murillo-Sánchez, and D. Gan, "MATPOWER: A MATLAB power system simulation package," *Manual, Power Systems Engineering Research Center, Ithaca NY*, vol. 1, 1997.

[25] "Power Systems Test Case Archive." [Online]. Available: https://www.maths.ed.ac.uk/optenergy/NetworkData/howtouse.html

[26] K. L. Lange, R. J. Little, and J. M. Taylor, "Robust statistical modeling using the t distribution," *Journal of the American Statistical Association*, vol. 84, no. 408, pp. 881–896, 1989.

[27] H. Tak, J. A. Ellis, and S. K. Ghosh, "Robust and Accurate Inference via a Mixture of Gaussian and Student'st Errors," *Journal of Computational and Graphical Statistics*, vol. 28, no. 2, pp. 415–426, 2019.

[28] M. Vallisneri and R. van Haasteren, "Taming outliers in pulsar-timing data sets with hierarchical likelihoods and Hamiltonian sampling," *Monthly Notices of the Royal Astronomical Society*, vol. 466, no. 4, pp. 4954–4959, 2017.

[29] M. D. Hoffman and A. Gelman, "The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1593–1623, 2014.

[30] "Brain Connectivity Toolbox." [Online]. Available: https://sites.google.com/site/bctnet/

[31] M. Waskom and the seaborn development team, "mwaskom/seaborn," *Zenodo*. [Online]. Available: https://doi.org/10.5281/zenodo.592845

[32] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in Python using PyMC3," *PeerJ Computer Science*, vol. 2, p. e55, 2016.

[33] A. Gelman, "Scaling regression inputs by dividing by two standard deviations," *Statistics in Medicine*, vol. 27, no. 15, pp. 2865–2873, 2008.

[34] B. P. Carlin and T. A. Louis, *Bayesian Methods for Data Analysis, Third Edition.* . Chapman and Hall/CRC, 2008.

[35] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*. CRC press, 2013.