Node Type Distribution and Its Impacts on Performance of Power Grids

Fei Xue, Shaofeng Lu, Ettore Bompard, Ciwei Gao, Lin Jiang, Xiaoliang Wang

***Abstract*—The theory of complex networks has been studied extensively since its inception. However, until now, the impact of the node-type distributions is related to network topology and cannot be evaluated independently. In this paper, a network structure is modeled via an adjacency matrix (network topology) and a set of node type distribution vectors. Three specific issues that need to be considered for node type distributions in smart grid testing and planning are summarized in this work. First, a set of metrics are proposed and defined to evaluate the impact of node type distributions on network performance independently. Second, another metric named the generation distribution factor is proposed to evaluate the distribution of generation buses resulting from the specific function and purpose of power grids and by considering the distribution of load buses as given conditions. Third, another metric, i.e., the power supply redundancy metric based on entropy, is proposed to evaluate the inequality of load in power supply. Finally, a discrimination factor is defined to ensure the overall evaluation and comparison of different networks is made for this inequality. All proposed metrics can be applied to IEEE-30, IEEE-118, IEEE-300 bus systems as well as Italian power grid components. The simulation results indicate that the IEEE-118 system has best node type distribution and minimum discrimination; the Italian system has the worst node-type distribution and most serious discrimination of load power supply.**

***Index Terms*—Complex Network, Node Type Distribution, Network Structure, Power Supply Redundancy.**

NOMENCLATURE

i, j	Bus index.
$g(d)$	Generation (load) bus.
G	Set of generator buses.
D	Set of load buses.
A	Adjacency matrix.
N	Network with a set (that contains all nodes), whose dimension is *N*.
N_G	Number of generator buses.
N_D	Number of loads buses.
N_{GD}	Number of nodes as both a generation bus and a load bus.
Y	Network model.
R	Benchmark network.
z_{ij}	Impedance of the line connecting between i and j.
Z_g^d	Equivalent impedance between g and d.
z_{gg}, z_{gd}, z_{dd}	Corresponding elements in the impedance matrix.
C_g^d	Power transmission capacity between g and d.
P_{ij}	Capacity (power flow limit) of the line connecting between i and j.
P_l^{max}	Power flow limit of line *l*.
f_l^{gd}	PTDF for line *l* when transferring power from g to d.
A_{ij}	Element of adjacency matrix $\boldsymbol{A}$.
$\boldsymbol{U}_S(\boldsymbol{U}_D)$	*N*-dimension source (sink) node distribution vector.
$U_i^S(U_i^D)$	Each element in $\boldsymbol{U}_S$ ($\boldsymbol{U}_D$).
$\boldsymbol{A}_B(\boldsymbol{A}_P)$	Extended network topology.
$E(\mathbf{Y})(E(\mathbf{R}))$	Net-ability of **Y** (**R**).
$ND(\mathbf{Y})$	Node type distribution factor *ND* for network **Y**.
D_R	Average distance of benchmark network **R**.
D_Y^{gd}	Average distance of power grid **Y**.
$D_Y^{gg}(D_Y^{dd})$	Average distance between any two generation (load) buses.
$NDes(\mathbf{Y})$	Nodes distribution density for power grid **Y**.
PS_g^d	Power supply scheme.
$EPS(PS_g^d)$	Performance index for a specific power supply scheme.
$ENode(g)$	Sum of performance from g to all load buses in $\boldsymbol{U}_D$.
$E(PS_d)$	Sum of performance from all generation nodes to *d*.
$GDFM(\mathbf{Y})$	Generation distribution factor metric of **Y**.
p_{gd}	Weight of performance index from a specific power source g to a specific load d.
PR_d	Power supply redundancy for the specific load bus *d*.
$Ave(PR_d)$	Average power supply redundancy for all load buses.
$DF(\mathbf{Y})$	Discrimination factor for network **Y**.

This work was supported in part by the Research Development Fund (RDF-15-02-14) of Xi'an Jiaotong-Liverpool University, in part by the research project "Study on Key Technologies for Energy-Internet-Oriented Microgrids" of SDIC Baiyin Wind Power Co., Ltd..

Fei Xue, Shaofeng Lu and Xiaoliang Wang are with the Department of Electrical and Electronic Engineering, Xi'an Jiaotong-Liverpool University, No. 111 Ren'ai Road, Suzhou Industrial Park, Suzhou, P.R. China (phone: +86 512 8816 1409; e-mail: Fei.Xue@xjtlu.edu.cn).

Ettore Bompard is with Politecnico di Torino in Italy.

Ciwei Gao is with Southeast University, Nanjing, P.R. China.

Lin Jiang is with the University of Liverpool in United Kingdom.

I. Introduction

AFTER examining studies on the significance of small-world and scale-free characteristics in complex network models [1][2], complex networks have been considered a promising direction when analyzing and evaluating networking issues and networked infrastructure systems. Network performance may be greatly influenced by corresponding network structures [3]-[7]. Furthermore, to estimate the average performance of a network, a concept of global efficiency was proposed by considering the average distance of the shortest path between any couple of nodes. This was then further applied for some typical networked systems [8]-[12].

However, in all pure topological models, every element of each node and edge are considered unweighted and non-directional. Therefore, many heterogeneous factors, such as line capacity and impedance, are neglected by pure topological analysis; the heterogeneous environment may have called a different result and may have a negative impact on the corresponding results. Therefore, a set of new metrics including net-ability [13][14], entropic degree [15] and electrical betweenness [16] were proposed for power grids extended topological models by taking into account these heterogeneous factors.

Recently, complex networks have been further applied for testing or planning Smart Grids. In [17], researchers proposed a method that automatically generates testing networks for smart grids. The influence of impedance distribution was considered in this model, but the distributions of node types were not addressed. In [18], a topological planning tool for upgrading conventional distribution networks has been put forward. In their model, a random sample of the nodes in the network (40% of the nodes whose half represents source nodes and the other half represents destination nodes) was evaluated. Therefore, the node type distribution was considered a random factor that was not an influential element to be specially analyzed. The network topology of South Korea power grids was analyzed in [19], only distribution of load nodes was generally discussed, no specific method was suggested. On the contrary, allocation of generation nodes was discussed in [20] with given network topology and load nodes. But the definition of community modularity and cascading failure models were not consistent with the physical rules of the power system. In [21], the generation/load layout was considered an independent topological factor. The resulting analysis emphasizes on the comparison of centralized or distributed power generation; but no appropriate metric was proposed. In [22][23], the nodes can be classified into source nodes and sink nodes, but their impact of distribution was not assessed independently.

Therefore, in summary, in the most existing research studies, node type distributions were considered inseparable from the network topology. However, due to constraints in investment and land supply, existing network topology structures are not possible to be significantly extended or changed. But the loads and energy sources, especially renewable energy sources, are expected highly developed in the future. Therefore, the generation/load layout may be an important but independent factor to be considered by the planners, and promising for real applications in smart grid planning and testing. Since the distributions were correlated with evaluation of network topologies, and has not been effectively analyzed as independent factor; this event can happen in both power grids and other complex network fields. Consequently, this paper is aimed at unraveling it from network topology and evaluates its independent impact on network performance. We summarize three major issues for the node type distribution:

(1). To evaluate the impact of generation/load layout that is independent from network topology. This issue has been discussed but not appropriately addressed in [21].

(2). To ensure optimal siting of generation nodes with given network topology and distribution loads. This issue was considered, but not fully justified in [20].

(3). To analyze the impact of load node distribution on power supply performance, the issue was partly mentioned, but not comprehensively modeled in [19].

In this paper, we address these three main issues by three types of metrics as a comprehensive analyzing framework. In section II, a network structure is defined via a network topology (adjacency matrix) and node type distribution vectors. The two methods are related by independent factors. In section III, a set of metrics are proposed and defined for estimating the independent impacts of generation/load layout on network performance. Moreover, in section IV, the load node distributions are given as conditions and another metric named the generation distribution factor metric (GDFM) is proposed to evaluate the distribution of generation nodes. In section V, when generation node distribution is provided, the load nodes will have unequal power supply service and security. Therefore, another metric named the power supply redundancy, based on entropy, is proposed to evaluate this inequality. A brand-new concept of load discrimination is proposed. In section VI, all proposed metrics are tested by the IEEE-30, IEEE-118, IEEE-300 bus systems and an Italian power grid. Simulation results are shown and discussed in this section. Conclusions are drawn in section VII.

II. Network Topology and Nodes Type Distribution

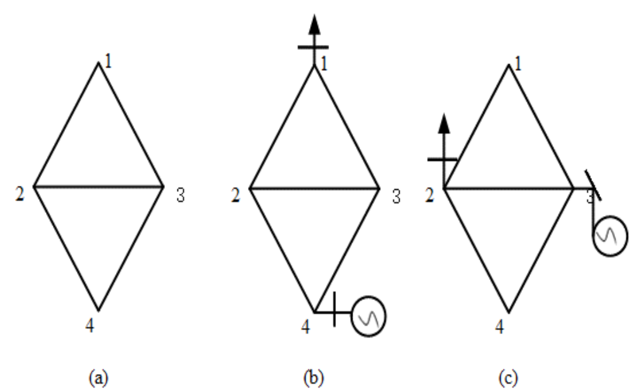


Fig. 1 Same network topology with different node type distribution

Figure 1 is an intuitive explanation of network topology and node type distribution. Case (a) is the network topology in terms of interconnecting relations between 4 nodes and 5 lines. Case (b) and (c) are examples of different node type distributions (generation and load at 4 and 1, or at 3 and 2) with the same network topology. In previous studies, node type distributions were considered inseparable from network topology. They were correlated in the model definition and evaluated in both power grids and other complex network fields. No specific method to distinguish node type distribution based on the network topology was ever proposed and no specific analyzing method to quantitatively evaluate its independent impact was ever developed (according to our knowledge). Therefore, in this section, we first propose the formal definition of node type distribution that is different from topology. This could be applied to power grids, but can also promote further studies in other similar fields.

In network science, a network topology could be characterized by a corresponding adjacency matrix **A** [24]. For an element A_{ij} of the adjacency matrix, we have:

$$A_{ij} = \begin{cases} 1, & \text{an edge going from node } i \text{ to node } j; \\ 0, & \text{otherwise;} \end{cases} \tag{1}$$

In a weighted network, A_{ij} could be the weight of the corresponding connection via the edge [25]

In a complex network, a node with zero input flow is called the source node and a node with zero output flow is the called sink node [22][23]. In pure topological analysis [1][2], any node in a network is considered as both a source and a sink node. This could be true for some networks, such as social networks or transportation networks. But it may not be true for some physical networks, such as power grids, where source nodes and sink nodes are just two subsets of all nodes.

Therefore, for a network with a set **N** (that contains all nodes), whose dimension is N, an N-dimension source node distribution vector can be defined as:

$$\boldsymbol{U_S} = [U_1,\ U_2, \dots, U_N]^T \tag{2}$$

For each element in $\boldsymbol{U_S}$, we have:

$$U_i^S = \begin{cases} 1, & \text{node } i \text{ is a source node;} \\ 0, & \text{otherwise;} \end{cases} \tag{3}$$

An N-dimensional sink node distribution vector is defined by

$$\boldsymbol{U_D} = [U_1,\qquad U_2, \dots, U_N]^T \tag{4}$$

Therefore, for each element in $\boldsymbol{U_D}$ we have

$$U_i^D = \begin{cases} 1, & \text{node } i \text{ is a sink node;} \\ 0, & \text{otherwise;} \end{cases} \tag{5}$$

Then a network model **Y** could be represented as

$$\mathbf{Y} = \{\mathbf{A}, \boldsymbol{U_S}, \boldsymbol{U_D}\} \tag{6}$$

A characterizes the network topology, while $\boldsymbol{U_S}$ and $\boldsymbol{U_D}$ indicate the node type distribution. The model, in pure topological analysis terms, can be found in [1][2] and is a special case for equation (6), where all elements in $\boldsymbol{U_S}$ are 1 and all elements in $\boldsymbol{U_D}$ are 1. For such a network, we can call it a homogenous network; otherwise, where only subsets of **N** are source or sink nodes, we can call it a heterogeneous network.

To construct a weighted network model for power grids, we need to consider what physical features are considered as the connection weights in the adjacency matrix. However, there are two features of the transmission line that are related to power transmission, i.e., impedance can be used to describe electrical distance and power transmission capacity [13]-[16]. So, for power grids, we can construct two adjacency matrices $\boldsymbol{A_B}$ and $\boldsymbol{A_P}$.

For each element in $\boldsymbol{A_B}$ we have

$$A_{ij}^B = \begin{cases} \frac{1}{z_{ij}}, & \text{a line connecting } i \text{ and } j; \\ 0, & \text{otherwise;} \end{cases} \tag{7}$$

where z_{ij} is the impedance of the line connecting between i and j.

For each element in $\boldsymbol{A_P}$ we have

$$A_{ij}^P = \begin{cases} P_{ij}, & \text{a line connecting } i \text{ and } j; \\ 0, & \text{otherwise;} \end{cases} \tag{8}$$

where P_{ij} is the capacity (power flow limit) of the line connecting between i and j.

For each element in $\boldsymbol{U_S}$, U_i^S is 1 if node i is a generation node. For each element in $\boldsymbol{U_D}$, U_i^D is 1 if node i is a load node.

Therefore, a power grid **Y** can be indicated as:

$$\mathbf{Y} = \{\mathbf{A}_B, \mathbf{A}_P, \mathbf{U}_S, \mathbf{U}_D\} \tag{9}$$

$\boldsymbol{A_B}$ and $\boldsymbol{A_P}$ represent the extended network topology; $\boldsymbol{U_S}$ and $\boldsymbol{U_D}$ represent the node type distribution.

III. Node Type Distribution Factor and Distance Distribution

This section is aimed at evaluating the impact of the generation/load layout space, which is independent from network topology. To evaluate the performance of a network, Global Efficiency was first proposed to determine the pure topological analysis [26]. An updated concept is then defined as the net-ability that was proposed for the performance of power grids and by considering physical weights and special rules in electrical engineering [13][14]

$$E(\mathbf{Y}) = \frac{1}{N_G N_D} \sum_{g \in G} \sum_{d \in D} \frac{C_g^d}{Z_g^d} \tag{10}$$

G and D are the sets of generator buses and load buses, respectively. N_G is the number of generator buses and N_D is the number of loads buses. The power transmission capacity C_g^d is based on the power injection at generation bus g that is withdrawn at load bus d when the first line in all lines connecting g and d reaches its limit P_l^{max} [13][14]:

$$C_g^d = \min_{l \in L} \left(\frac{P_l^{max}}{\left| f_l^{gd} \right|} \right) \tag{11}$$

where f_l^{gd} is the Power Transfer and Distribution Factor (PTDF) for line l when transferring power from g to d. According to the electrical circuit theory, the equivalent impedance Z_g^d can be expressed as [13][14]:

$$Z_g^d = z_{gg} - 2z_{gd} + z_{dd} \tag{12}$$

where z_{gg}, z_{gd} and z_{dd} are corresponding elements in the impedance matrix (inverse matrix of admittance matrix) of the network.

To quantitatively evaluate the impact of generation/load layout on performance of network **Y** indicated by (9), we firstly need to construct a benchmark network **R**:

$$\mathbf{R} = \{\mathbf{A}_B, \mathbf{A}_P, \mathbf{U}_S^R, \mathbf{U}_D^R\} \tag{13}$$

where all elements in $\mathbf{U}_S^R$ are 1 and all elements in $\mathbf{U}_D^R$ are 1. That means any bus is both a generation bus and a load bus. Then the network **Y** in equation (9) has the same network topology with **R** represented by the same $\mathbf{A}_B$ and $\mathbf{A}_P$, but they have different generation/load layout indicated by $\boldsymbol{U_S}$ and $\boldsymbol{U_D}$.

Then a node type distribution factor ND for network **Y** can be defined as the relation of net-ability between **Y** and **R**:

$$ND(\mathbf{Y}) = E(\mathbf{Y}) / E(\mathbf{R}) \tag{14}$$

Because the network topology ($\mathbf{A}_B$ and $\mathbf{A}_P$) of **Y** and **R** are identical, the difference of performance between them is completely caused by generation/load layout. **Any node in R is**

both a generation and a load bus, that means both generation and load are fully distributed in the network. So, network R could be considered as a benchmark. *ND* could be a factor to indicate how the performance affected by the generation/load layout in **Y**. Even with two networks **Y1** and **Y2**, whose network topologies and scales may be totally different, by comparing with corresponding different benchmarks **R1** and **R2** respectively, *ND*(**Y1**) and *ND*(**Y2**) can still be directly compared for their different extents of impact from generation/load layout.

Alternatively, the impact of different generation/load layout can also be reflected by variation in distance features. In a pure topological analysis, the concept of distance is defined as the length of the shortest path between two nodes. The average path length of a network is the average value of distances between any pair of nodes in the network. The reciprocal of distance was the essential element in defining Global Efficiency as network performance [26]. In power grids, an electrical distance was defined as the equivalent impedance between a generation bus and a load bus as shown by equation (12) in [13][14].

In this paper, the following definitions for different average distances are made to study the impact of generation/load layout on the distance features. First, the average distance of benchmark network **R** is defined as:

$$D_R = \frac{1}{N(N-1)} \sum_{i,j\in \mathbf{N}, i\neq j} Z_i^j \qquad (15)$$

The average distance of power grid **Y** is defined as:

$$D_Y^{gd} = \frac{1}{N_G N_D} \sum_{g\in G} \sum_{d\in D} Z_g^d \qquad (16)$$

The average distance between any two generation buses and any two load buses are respectively defined as:

$$D_Y^{gg} = \frac{1}{N_G(N_G-1)} \sum_{i,j\in G, i\neq j} Z_i^j \qquad (17)$$

$$D_Y^{dd} = \frac{1}{N_d(N_d-1)} \sum_{i,j\in D, i\neq j} Z_i^j \qquad (18)$$

The relationship and the local time these distance distributions can be used to study the features of distance affected by node type distribution. For example, the following relations and corresponding impacts can be found:

a). $D_Y^{gd} \gg D_R$, The average distance of **Y** is much larger than that of benchmark **R**. Compared with average distance of original network topology in **R**, the generators are much further from loads due to generation/load layout increasing the distance and cost in power transmission.

b). $D_Y^{gd} \gg D_Y^{gg}$ and $D_Y^{gd} \gg D_Y^{dd}$, the distribution of generation and load buses are quite uneven. The generation buses concentrate as a community and load buses concentrate as a community respectively, but they are far from each other. This is not efficient for power transmission and may be vulnerable for failures between them.

c). $D_Y^{gg} \approx D_Y^{gd} \approx D_Y^{dd} \ll D_R$, generation buses and load buses all concentrate to a small fraction of the power grid. Although the performance of power transmission may be efficient, a large part of the network may be not fully utilized.

The above-mentioned cases are just some examples; different observations can be made based on specific relations.

Meanwhile, as not all nodes in **Y** may be the generation buses or load buses, a nodes distribution density for power grid **Y** can be defined as:

$$NDes(\mathbf{Y}) = {(N_G N_D - N_{GD})}\Big/{N(N-1)} \qquad (19)$$

where N_{GD} is the number of nodes as both a generation bus and a load bus. This density is to indicate how many generation-load pairs compared with the maximum possible number in **R**. $NDes(\mathbf{Y})$ is to compare the power transmission scale and the network scale. For example, in an extreme example with a very large network, there are only one generation bus and one load bus. Then the nodes distribution density is extremely low which means the network scale is not fully utilized and inefficient. In opposite, for network **R** where any bus is both a generation bus and a load bus, the *NDes* has the highest value as 1. In designing a power grid, a reasonable density should be considered.

IV. Generation Buses Distribution

The siting of the generation nodes were addressed in [20] via complex network approaches. It is undoubted that this is a critical issue in smart grid planning and testing. But some assumptions are not quite consistent with engineering conditions and rules in [20], such as the line current capacity and the detection of community structure. This section will select top candidate generation buses from network performance perspective or quantify the generation buses distribution compared with a benchmark.

To perform this, we firstly need to define a power supply scheme using a quadruplet as follows:

$$PS_g^d = \{g, d, C_g^d, Z_g^d\} \qquad (20)$$

A power supply scheme is just a scenario that transmit power from a generation bus *g* to a load bus *d*. It is related to the concrete positions of *g* and *d*, the maximum transmission capacity C_g^d between them, and the electrical distance Z_g^d between them. Following the idea of net-ability [13][14], the performance index for a specific power supply scheme can be defined as:

$$EPS(PS_g^d) = \frac{C_g^d}{Z_g^d} \qquad (21)$$

When considering positions of load buses in $\boldsymbol{U}_D$ as a given condition, for any bus $g \in \mathbf{N}$ in the network, the power supply performance index of bus g is defined as the sum of performance from g to all load buses in $\boldsymbol{U}_D$:

$$ENode(g) = \sum_{d\in D} \frac{C_g^d}{Z_g^d} \qquad (22)$$

To evaluate the generation buses distribution of network **Y**= {$\mathbf{A}_B$, $\mathbf{A}_P$, $\mathbf{U}_S$, $\mathbf{U}_D$}, we consider the load bus set *D* with their positions in $\boldsymbol{U}_D$ and N_G (dimension of generation bus set *G*) as fixed constraints. The solution is to find the top N_G candidate buses for generation based on power supply performance. For all buses in set **N**, we can find the bus corresponding to maximum power supply performance index:

$$g = \operatorname*{argmax}_{g\in\mathbf{N}} \sum_{d\in D} \frac{C_g^d}{Z_g^d} \qquad (23)$$

Following that, the identified bus g is removed from **N**. Then this process will be repeated for N_G times to find the top N_G buses.

Compared with $\mathbf{Y}=\{\mathbf{A}_B, \mathbf{A}_P, \mathbf{U}_S, \mathbf{U}_D\}$, another benchmark network can be constructed as $\mathbf{R}_g=\{\mathbf{A}_B, \mathbf{A}_P, \mathbf{U}_S^{Rg}, \mathbf{U}_D\}$. In $\mathbf{U}_S^{Rg}$, only the elements corresponding to the top $\boldsymbol{N}_G$ buses are 1. Then the network topology $(\mathbf{A}_B, \mathbf{A}_P)$, load buses distribution $(\mathbf{U}_D)$ and number of generation buses N_G for $\mathbf{Y}$ and $\mathbf{R}_g$ are the same. In $\mathbf{R}_g$, the top N_G buses are selected as generation buses. Therefore, the generation distribution factor metric (GDFM) of $\mathbf{Y}$ can be defined as relation of net-ability between $\mathbf{Y}$ and $\mathbf{R}_g$:

$$GDFM(\mathbf{Y}) = E(\mathbf{Y}) / E(\mathbf{R}_g) \tag{24}$$

V. Power supply redundancy Metric

Corresponding to the third issue mentioned in introduction, this section is aimed to analyze power supply performance for load nodes with network topology and generation nodes as given conditions.

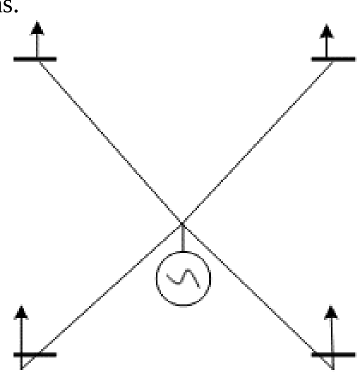


Fig. 2 Vulnerable power supply with single generator for multiple loads

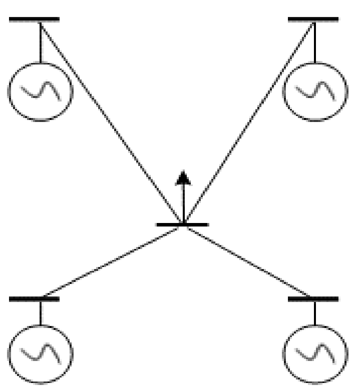


Fig. 3 Redundant power supply for single load with multiple generators

Reference [19] has concluded that loads need to be homogeneously distributed and decentralized to reduce vulnerability of power grids. However, the relative distribution of generation nodes regarding load nodes was not comprehensively considered and the unequal position of each single load node in power supply was not evaluated.

Considering figure 2 and figure 3, the capacities and impedances of all lines in these two networks are the same. By reviewing the definition of net-ability in equation (10), the total number of power supply schemes (or the total number of generation and load pairs) are the same for the two networks. **Therefore, the network performances evaluated by net-ability for these two networks are the same.**

However, it is easy to have an intuitive impression that **the case in figure 3 is more reliable than that in figure 2**. In figure 2, all load buses depend on only one generation bus with no other backup for power supply. Therefore, if the generation bus fails, all these loads will be affected. However, in figure 3, the load has four possible power sources. The failure of one or even more generation buses may not interrupt its power supply as it has a large redundancy.

In figure 4, in both case (a) and (b), the load has two possible generation buses for power supply. The weight of performance index defined in (21) for two power supply schemes are shown in the figure (0.5 and 0.5, 0.1 and 0.9). However, these weights between two power supply schemes in two cases are very different. In case (a), both weights are the same and the power supply to the load depends equally on these two generation buses. But in case (b), the weight for one generation bus is 0.9 and much larger than the other one being 0.1. The power supply to the load bus greatly depends on one generation bus. If this bus fails, the load bus will be much affected. So, power supply redundancy depends on both the number of power sources and the distribution of weights.

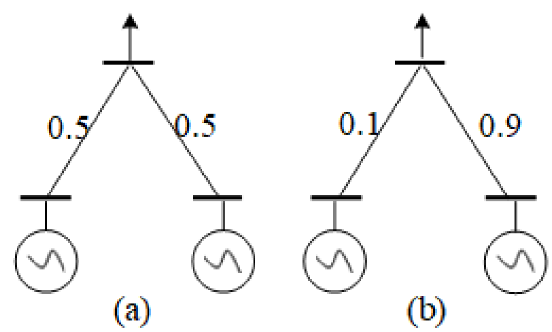


Fig. 4 Different entropy of power supply by different weight distribution

From the above discussions, we can see that for one specific load bus, its concrete distribution relations with other generation buses in the network may seriously influence the power supply redundancy for it. To evaluate this, we resort to the concept of entropy to analyze the corresponding redundancy. Entropy has been applied to evaluate the redundancy of paths between a pair of generation and load in [14]. Here we use it to evaluate redundancy in power supply for a specific load bus regarding all generation buses. The entropy for case (a) in figure 4 is higher than case (b), this is consistent with the concept of redundancy as we discussed.

By reconsidering the performance of power supply scheme in equation (21), we can further define the total power supply performance index for a specific load bus d as the sum of performance from all generation nodes to d:

$$E(PS_d) = \sum_{g \in G} \frac{C_g^d}{Z_g^d} \tag{25}$$

The weight of performance index from a specific power source g to a specific load d can be calculated as:

$$p_{gd} = \frac{E(PS_g^d)}{E(PS_d)} \tag{26}$$

The power supply redundancy for the specific load bus d based on concept of entropy [13][14] can be defined as:

$$PR_d = \left(1 - \sum_{g\in G} p_{gd}\log p_{gd}\right)\sum_{g\in G}\frac{C_g^d}{Z_g^d} \quad (27)$$

If we select 10 as the base of logarithm in (27), in figure 2, for any one load bus, there is only one generator connected, so $p_{gd} = 1$. In the case of figure 3, the load buses relate to four generators, for each one we have $p_{gd} = 0.25$. According to the definition of entropy we have $[4 \times (-0.25\log_{10}0.25)] > (-1\log_{10}1)$. For cases in Figure 4, we have $[(-0.5log_{10}0.5) + (-0.5log_{10}0.5)] > [(-0.1log_{10}0.1) + (-0.9log_{10}0.9)]$.

In principle, the customers in the same power grid should have equal rights for power supply service in terms of quality and security. However, due to their different distribution features regarding generation buses, discrimination must exist in realities. By comparing power supply redundancies of different load buses, we can evaluate if the load buses in the same network have very different power supply service quality and security. From the perspective of the entire network, a discrimination factor can be defined. First, the average power supply redundancy for all load buses is calculated as:

$$Ave(PR_d) = \sum_{d\in D} PR_d \Big/ N_D \quad (28)$$

Then the discrimination factor for network **Y** can be defined by standard deviation as:

$$DF(\mathbf{Y}) = \sqrt[2]{\sum_{d\in D}\left[\frac{PR_d - Ave(PR_d)}{Ave(PR_d)}\right]^2 \Big/ N_D} \quad (29)$$

VI. CASE STUDY

To verify the proposed metrics, we select four systems with different scales from small to larger, i.e. the IEEE-30, IEEE-118, IEEE-300 bus systems and an Italian power grid [15]. The results for node type distribution factor and distance distribution are summarized in table I. All results are indicated by per unit values.

TABLE I
NODES DISTRIBUTION FACTOR AND DISTANCE DISTRIBUTION

System	ND	D_R	D_Y^{gd}	D_Y^{gg}
IEEE-30	0.9304	0.3445	0.3444	0.2916
IEEE-118	1.0206	0.2137	0.2155	0.2139
IEEE-300	0.8662	0.7089	0.6025	0.2566
ITALIAN	0.4278	0.0583	0.0598	0.0684

To clarify the explanation, we take the IEEE-30 bus system as an example which is simple to observe and understand. For IEEE-30, we have $D_Y^{gd} \approx D_R$, that means the node type distribution does not obviously change the distance features. By considering ND(IEEE-30) = 0.9304, the generation/load layout does not influence the network performance regarding the original network topology. However, with $D_Y^{gd} \approx D_R \gg D_Y^{gg}$, the average distance between generation buses is much smaller than the average distance of the whole network or the average distance between generation and load buses. The structure of the IEEE-30 bus system is shown in figure 5. It is obvious that the whole network can be divided into two communities. And most generation buses concentrate in the top community with close distance, this is consistent with what mentioned above.

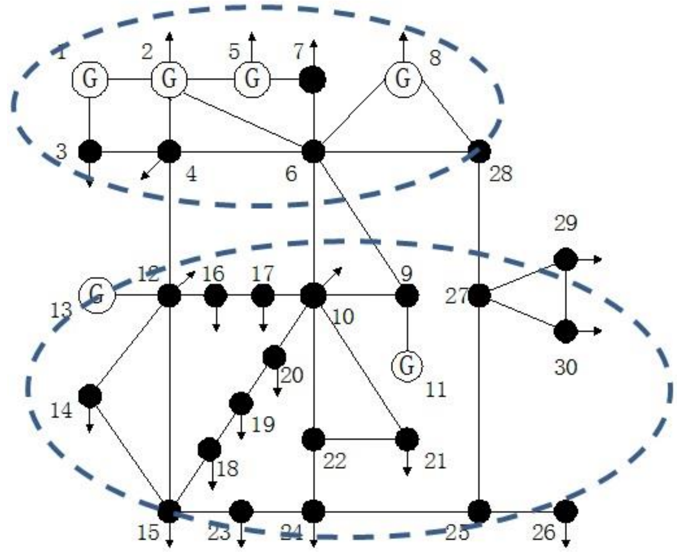


Fig. 5 The tested IEEE30-bus system with top and bottom communities

From overall perspective, among these 4 testing systems, the IEEE-118 bus system has the best generation/load layout. Its node type distribution factor ND(IEEE-118) = 1.0206 indicates that the generation/load layout even improve the network performance regarding original network topology. And with $D_Y^{gg} \approx D_Y^{gd} \approx D_Y^{dd} \approx D_R$, the distance distribution is quite even in the whole network. The Italian power grid has the lowest node type distribution factor ND(ITALIAN)=0.4278, that means its generation/load layout worsen the network performance seriously regarding its original network topology.

Table II indicates the generation distribution, nodes distribution density and the discrimination factor of these tested systems. For better comparison, the node type distribution factors are also shown in that table.

TABLE II
GENERATION DISTRIBUTION AND DISCRIMINATION FACTORS

System	ND	$GDFM$	$NDes$	DF
IEEE-30	0.9304	0.5634	0.1775	0.5725
IEEE-30+	1.6513	1	0.1775	0.4713
IEEE-118	1.0206	0.8214	0.0798	0.2969
IEEE-300	0.8662	0.5157	0.0432	0.5550
ITALIAN	0.4278	0.1773	0.0238	0.6624

We still take the IEEE-30 bus system as an example. There are totally 6 generation buses in the system. Their bus IDs are 1, 2, 5, 8, 11, 13. Then by equation (23), the top 6 buses for generation siting are found as bus 6, 4, 22, 12, 10, 2. If the generators of the IEEE-30 bus system are moved to these 6 top buses, then the corresponding benchmark system $\mathbf{R}_g$ in equation (24) is got. This system is denoted as IEEE-30+ in table II. It is observed that by improving generation distribution, the node type distribution factor ND is greatly increased compared with the IEEE-30 bus system. And furthermore, compared with the IEEE-30 bus system, the discrimination factor of IEEE-30+ is obviously reduced by redistribution of generation buses. From figure 5, we can see that the generation buses concentrate at the top community in IEEE-30 (communities are indicated by dashed circles). The load buses in the top community of course may have better power supply redundancy and reliability. But the load buses at bottom community may not have equal service with similar extent of reliability because they are relatively far from the power sources. The failures of lines at the border between the two communities may seriously threaten the power supply to the loads in the bottom community. However, the generation buses are more evenly distributed in the whole system in IEEE-30+. So, the power supply redundancy and reliability can be remarkably improved for all load buses in terms of high power supply redundancy and low discrimination factor.

Based on the evaluation listed 2 in Table I and Table II, among the 4 tested systems, it is obvious that the IEEE-118 system has the best generation distribution with a high *GDFM*, but the Italian system has a much lower *GDFM* corresponding to worst generation distribution. Correspondingly, the IEEE-118 system has the smallest discrimination factor *DF* due to its better generation distribution, and the Italian system has a much larger *DF* because of its poor generation distribution. It is concluded that among the four tested systems, the IEEE-118 bus system has the best node type distribution, and the Italian power grid has the worst case. Following, this result will be justified and analyzed **by community structure, betweenness distribution and power congestion simulation**.

Firstly, according to the method in [27], the IEEE-118 system was partitioned into 3 communities (as shown in figure 6) and the Italian power grid was partitioned into 6 communities by electrical coupling strength. The ratio between numbers of generation buses and load buses (N_G^C/N_D^C for community c) in each community was calculated. The results for the IEEE-118 system are:

$$N_G^1/N_D^1 = 0.7727$$
$$N_G^2/N_D^2 = 0.8696$$
$$N_G^3/N_D^3 = 0.8947$$

The results for the Italian power grid are:

$$N_G^1/N_D^1 = 0.4861$$
$$N_G^2/N_D^2 = 0.3571$$
$$N_G^3/N_D^3 = 0.7021$$
$$N_G^4/N_D^4 = 0.3621$$
$$N_G^5/N_D^5 = 0.4200$$
$$N_G^6/N_D^6 = 0.3600$$

It can be observed that the ratios of communities in IEEE-118 systems are quite similar, but those in the Italian system are quite different. This has proved that the generation/load layout in the IEEE-118 system is consistent with the tendency of community structures. However, the generation/load layout in the Italian system conflicts with the community structures.

Secondly, betweenness was widely used in assessing the responsibilities of concrete components in structural analysis of power grids [16][28]. From perspective of security, if all components have even distribution of betweenness, the system operation may not depend on any specific component seriously. If the node type distribution makes more uneven distribution of betweenness, the power supply security will be worsened. In figure 7, the red line indicates the betweenness distribution of the IEEE-118 system and the blue line is corresponding to its benchmark system R^{118} defined by equation (13). The horizontal axis is the ranking of line betweenness. For example, 1 on the horizontal axis indicates the line with the largest betweenness value. It can be observed that the betweenness distribution is much evener than its benchmark system. That means the node type distribution has improved the network structure considering security properties. However, figure 8 is the case for the Italian system. The tendency of betweenness distribution is not quite different from its benchmark system. The variance *Var* of betweenness distribution can be calculated for each system and compared with its benchmark system as:

$$Var(118)/Var(R^{118}) = 0.29$$
$$Var(Italian)/Var(R^{Italian}) = 0.62$$

Compared with the Italian system, the node type distribution of the IEEE-118 system has greatly improved the betweenness distribution with much more reduction of variance.

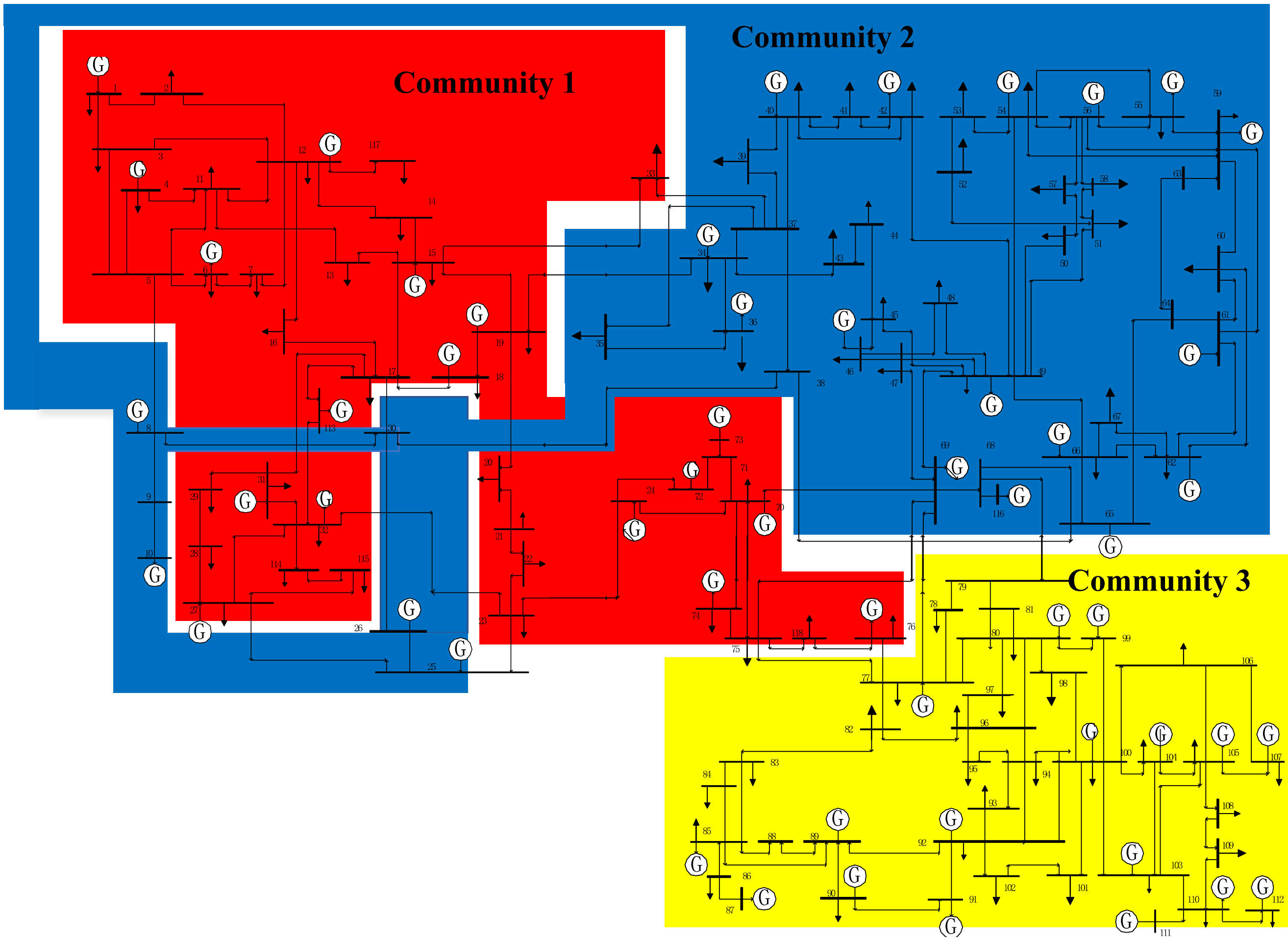
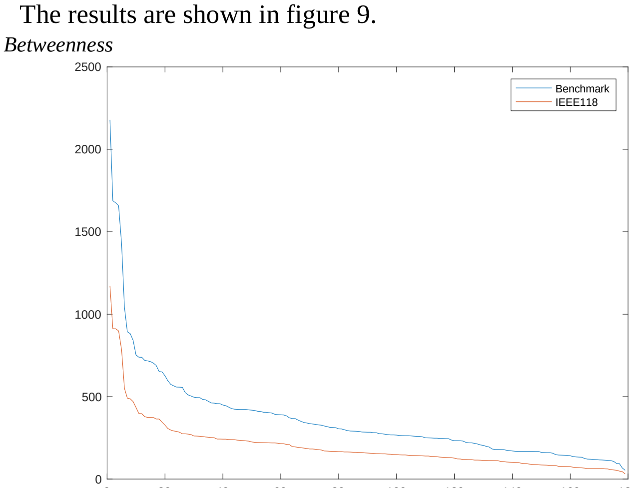


Fig. 6 The tested IEEE118-bus system with three communities

Thirdly, AC power flow models have been constructed based on MATPOWER for the IEEE30, IEEE118, IEEE300 and Italian systems. Top 5% transmission lines with largest betweenness values were removed step by step. Then the congested power $\sum_l (P^l - P^l_{max})/_D$ (only if $P^l > P^l_{max}$) has been calculated as the ratio between total power higher than the power flow limits P^l_{max} for each line and the system total load D.

The results are shown in figure 9.



Ranking

Fig. 7 Betweenness distribution of the IEEE-118 and benchmark system with same topology. The IEEE-118 system has more even distribution due to node type distribution.

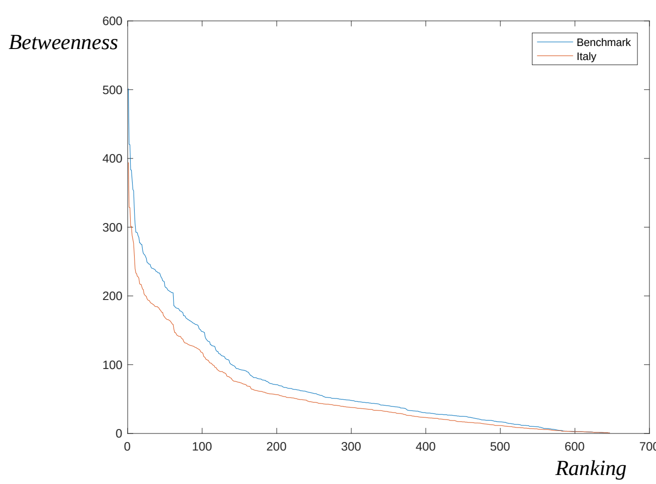


Fig. 8 Betweenness distribution of the Italian and benchmark system with same topology. Both have similar distribution due to no improvement from node type distribution.

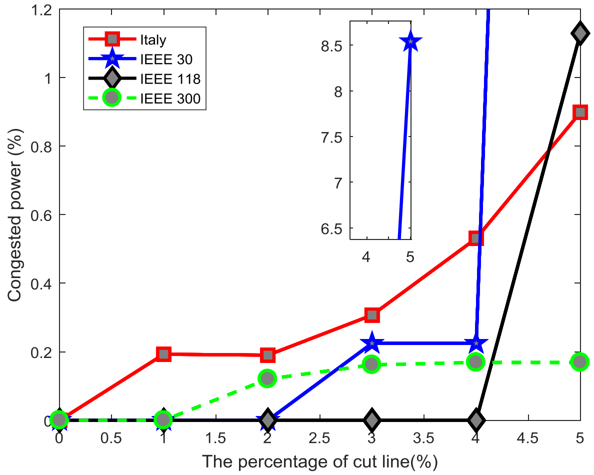


Fig. 9 Congested power regarding removed lines. The IEEE-118 system can withstand line failures more than the Italian system due to generation/load layout.

In figure 9, it can be observed that from 0% to 4% top lines removed, the IEEE118 system has no power flow higher than limit. But for the Italian system, with the same condition, much more power could be congested, and power supply to more loads could be impacted. The performances of the IEEE30 and IEEE300 systems are between the former two cases. The main reason is that the IEEE118 system has a better node type distribution, so it has better ability to withstand worsened network conditions. These are consistent with the results in TABLE I and TABLE II.

According to the evaluations from community structure, betweenness distribution and power congestion simulation, it has been justified that the IEEE-118 system has a much better node type distribution than the Italian system. This is consistent with the results from the metrics proposed in this paper.

VII. Conclusion

Previous work has applied the theory of complex networks to analyze networked infrastructure systems including power grids. Recently, complex networks are active in constructing testing models for smart grids or upgrading conventional power grids to smart grids. However, impacts of node types and their distribution have not been considered independently and comprehensively. This paper has summarized three related issues about node type distribution which have not been appropriately addressed in early studies. Then different methods and metrics are designed dedicated for these three issues. To our best knowledge, **it is the first time that node type distribution is defined and evaluated independently from network topology. And it is also the first time that the problem of load discrimination is proposed and evaluated.** All proposed metrics are from statistical perspective and consistent with dynamic simulation in congested power by removing critical lines. By comprehensive analysis in four power grids with different scales, especially for the IEEE118 and the Italian systems, the proposed metrics have been justified effective in analyzing and evaluating node type distribution. Taking the Italian system as an example, with further development of loads and new energy sources, it would be possible to improve the generation/load layout. Of course, the allocation of power sources depends on a lot of complicated factors, not only structure performance. But the proposed metrics could be integrated in the relevant decision-making process and make the planner aware of this perspective. **This can provide a new perspective and make insight in testing and planning models for smart grids.**

Furthermore, the value of this paper may not be limited to power grids because many other network systems may also have different types of nodes. **Similar metrics could be developed to construct effective analysis framework for other infrastructure systems.** In the future, node type distribution can be further studied for their impacts on structure features of cascading failures. And structural analysis for energy storage performance according to generation and load spatial distribution can also be implemented.

References

[1] D. J. Watts, S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, pp. 440–442, June . 1998.

[2] A. L. Barabási, R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp.509–512, Oct. 1999

[3] M. Rosas-Casals, S. Valverde, and R. Solé, "Topological vulnerability of the European power grid under erros and attacks," *Int. J. Bifurc. Chaos*, vol. 17, no. 7, pp. 2465–2475, 2007.

[4] R. Albert, I. Albert, and G. L. Nakarado, "Structural vulnerability of the North American power grid," *Phy. Rev. E, vol.* 69, p. 025103(R), Feb. 2004.

[5] D. P. Chassin and C. Posse, "Evaluating North American electric grid reliability using the Barabasi-Albert network model," *Phys. A, Statist. Mech. Appl*, vol. 355, pp. 667–677, Apr. 2005

[6] V. Latora and M. Marchiori, "Vulnerability and protection of infrastructure networks," *Phys. Rev. E*, vol. 71, no. 1, p. 015103(R), Jan. 2005.

[7] P. Crucitti, V. Latora, and M. Marchiori, "Locating critical lines in high-voltage electrical power grids," *Fluctuation Noise Lett*, vol. 5, no. 2, pp. L201–L208, 2005.

[8] P. Crucitti, V. Latora, and M. Marchiori, "Model for cascading failures in complex networks," *Phys. Rev. E*, vol. 69, p. 045104(R), Apr. 2004.

[9] A. E. Motter and Y.-C. Lai, "Cascade-based attacks on complex networks," *Phys. Rev. E*, vol. 66, p. 065102, Dec. 2002.

[10] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda, "Error and attack tolerance of complex networks," *Phys. A, Statist. Mech. Appl*, vol. 340, no. 1, pp. 388–394, Sep. 2004.

[11] R. Kinney, P. Crucitti, R. Albert, and V. Latora, "Modeling cascading failures in the North American power grid," *Eur. Phys. J. B*, vol. 46, pp. 101–107, Aug. 2005.

[12] P. Crucittia, V. Latorab, and M. Marchioric, "A topological analysis of the Italian electric power grid," *Phys. A, Statist. Mech. Appl*, vol. 338, pp. 92–97, Jul. 2004.

[13] S. Arianos, E. Bompard, A. Carbone, and F. Xue, "Power grid vulnerability: A complex network approach," *Chaos*, vol. 19, no. 1, p. 013119, Mar. 2009.

[14] E. Bompard, R. Napoli, and F. Xue, "Extended topological approach for the assessment of structural vulnerability in transmission networks," *IET Gener. Transm. Distrib.*, vol. 4, no. 6, pp. 716–724, Jun. 2010.

[15] E. Bompard, R. Napoli, and F. Xue, "Analysis of structural vulnerabilities in power transmission grids," *Int. J. Critic. Infrastruct. Protect.*, vol. 2, pp.5-12, May. 2009

[16] E. Bompard, D. Wu, and F. Xue, "Structural vulnerability of power systems: A topological approach," *Electric Power Systems Research*, vol. 81, pp. 1334–1340, July 2011.

[17] Z. Wang, A. Scaglione and R. J. Thomas, "Generating Statistically Correct Random Topologies for Testing Smart Grid Communication and Control Networks," in *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 28-39, June 2010.

[18] G. A. Pagani, M. Aiello, "From the grid to the smart grid, topologically," *Phys. A, Statist. Mech. Appl*, vol. 449. pp. 160–175, May. 2016.

[19] D. H. Kim, D. A. Eisenberg, Y. H. Chun, and J. Park, “Network topology and resilience analysis of south korean power grid,” *Phys. A, Statist. Mech. Appl.*, vol. 465, pp. 13–24, Jan. 2017.
[20] W. Z. Guo, H. Wang, and Z. P. Wu, “Robustness analysis of complex networks with power decentralization strategy via flow-sensitive centrality against cascading failures,” *Phys. A, Statist. Mech. Appl*, vol. 494, pp. 186–199, Mar. 2018.
[21] A. Azzolin, L. Dueñas-Osorio , F. Cadini , and E. Zio, “Electrical and topological drivers of the cascading failure dynamics in power transmission networks,” *Reliability Engineering and System Safety*, vol. 175, pp. 196–206, Jul. 2018.
[22] Z. Y. Wang, D. J. Hill, G Chen, and Z. Y Dong, “Power system cascading risk assessment based on complex network theory,” *Phys. A, Statist. Mech. Appl*, vol. 482, pp. 532–543, Sep. 2017.
[23] A. Dwivedi and X. Yu, "A Maximum-Flow-Based Complex Network Approach for Power System Vulnerability Analysis," in *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 81-88, Feb. 2013.
[24] A. Clauset, M. E. J. Newman, and C. Moore, “Finding Community Structure in Very Large Networks,” *Phys. Rev. E*, vol. 70, p. 066111, Dec. 2004.
[25] M. E. J. Newman, “Analysis of weighted networks,” *Phys. Rev. E*, vol. 70, p. 056131, Nov. 2004.
[26] V. Latora and M. Marchiori, "Vulnerability and protection of infrastructure networks," Phys. Rev. E, vol. 71, p. 015103, Jan. 2005.
[27] X. T. Xu, F. Xue, S. F. Lu, H. Y. Zhu, L. Jiang, and B. Han, “Structural and Hierarchical Partitioning of Virtual Microgrids in Power Distribution Network,” *IEEE Systems Journal*, 2018
[28] J. Yan, H. He and Y. Sun, "Integrated Security Analysis on Cascading Failure in Complex Networks," in *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pp. 451-463, Mar 2014.