frontiers | Frontiers in Energy Research

ORIGINAL RESEARCH
published: 14 July 2022
doi: 10.3389/fenrg.2022.909611

Probabilistic Optimal Power Flow-Based Spectral Clustering Method Considering Variable Renewable Energy Sources

Juhwan Kim[1], Jaehyeong Lee[2], Sungwoo Kang[1], Sungchul Hwang[3], Minhan Yoon[4] and Gilsoo Jang[1]**

[1]School of Electrical Engineering, Korea University, Seoul, South Korea, [2]Power System Planning Department, Korea Electric Power Corporation, Naju, South Korea, [3]Department of Electrical Engineering, Sunchon National University, Suncheon, South Korea, [4]Department of Electrical Engineering, Kwangwoon University, Seoul, South Korea

Power system clustering is an effective method for realizing voltage control and preventing failure propagation. Various approaches are used for power system clustering. Graph-theory-based spectral clustering methods are widely used because they follow a simple approach with a short calculation time. However, spectral clustering methods can only be applied in system environments for which the power generation amount and load are known. Moreover, it is often impossible to sufficiently reflect the influence of volatile power sources (e.g., renewable energy sources) in the clustering. To this end, this study proposes a probabilistic spectral clustering algorithm applicable to a power system, including a photovoltaic (PV) model (for volatile energy sources) and a classification method (for neutral buses). The algorithm applies a clustering method that reflects the random outputs of PV sources, and the neutral buses can be reclassified via clustering to obtain optimal clustering results. The algorithm is verified through an IEEE 118-bus test system, including PV sources.

Keywords: hierarchical spectral clustering, electric power system, photovolataics, power system analysis, expansion

OPEN ACCESS

Edited by:
Xingpeng Li,
University of Houston, United States

Reviewed by:
Maneesh Kumar,
Yeshwantrao Chavan College of Engineering (YCCE), India
Martin P. Calasan,
University of Montenegro, Montenegro

****Correspondence:***
Minhan Yoon
myoon@kw.ac.kr
Gilsoo Jang
gjang@korea.ac.kr

Specialty section:
This article was submitted to Smart Grids, a section of the journal Frontiers in Energy Research

Received: *31 March 2022*
Accepted: *30 May 2022*
Published: *14 July 2022*

Citation:
Kim J, Lee J, Kang S, Hwang S, Yoon M and Jang G (2022) Probabilistic Optimal Power Flow-Based Spectral Clustering Method Considering Variable Renewable Energy Sources. Front. Energy Res. 10:909611. doi: 10.3389/fenrg.2022.909611

1 INTRODUCTION

New and renewable energy resources have been increasingly used worldwide owing to energy policies aimed at achieving carbon neutrality (Cauz et al., 2020; Li et al., 2021). From a numerical perspective, the accumulated capacity of global wind power increased by a factor of 36.7 (from 16.9 to 621.6 GW) between 2000 and 2019. Within the same time frame, photovoltaic (PV) capacity increased by a factor of 480.8, from 1.2 GW in 2000 to 590.3 GW in 2019 (IRENA, 2021). Owing to the variability in wind and PV generation, wind and PV are treated as variable renewable energy (VRE) sources. In one instance in Denmark (where high renewable penetration has been achieved), VRE generation exceeded demand for 845 h and recorded a high demand of 213%. Such situations have become more frequent since the first occurrence in 2015 (Holttinen et al., 2021). Increases in the uncertainty and variability of VRE sources are expected to pose challenges to the secure operation of modern power systems. Accordingly, probability analyses have been applied to incorporate the uncertainties from VRE sources and converter-interfaced generation (Wiget et al., 2014; Leeuwen and Moser, 2017). Novel

TABLE 1 | Comparison of clustering references.

Paper	Advantages	Disadvantages
Sarajpoor et al. (2021)	Proposes a time aggregation framework for choosing representative periods for studies that include both wind and load data	Inapplicable to clustering content for network segmentation
Chai et al. (2018)	1) Novel index based on electrical distance and voltage capability 2) Network partition based on Tabu search	Local voltage regulations are required in applications
Cao et al. (2021)	Undertakes network partitioning considering the voltage sensitivity (based on electrical distance) to achieve decentralized control	Cannot guarantee application in transmission system because of its application in the distribution system
Ma et al. (2021)	Proposes graph-theory-based network partitioning algorithm to realize decentralized detection with a faster response	Numerous grouping steps for network partitioning
Si et al. (2021)	Summarizes the concepts and general process in electric load clustering for smart grids	Does not cover specific application cases
von Luxburg, (2007)	Presents the most common spectral clustering algorithms and derives them from scratch via several different approaches	Does not cover the application of the power system
Sánchez-García et al. (2014)	Provides a thorough theoretical justification of the use of spectral clustering in power systems, including the results of our methodology for several test systems	Clustering process considering the variability of VRE sources is neglected
Tyuryukanov et al. (2018)	Proposes an approach based on the orthogonal transformation of spectral clustering to closely fit the axes of the canonical coordinate system	Clustering process considering the variability of VRE sources is neglected
Amini et al. (2020)	Improves hierarchical clustering such that the generator coherency constraint can be included in the clustering process	Clustering process considering the variability of VRE sources is neglected
Bialek and Vahidinasab, (2022)	Offers a graph-theoretic justification for tree-partitioning based on spectral clustering	Does not include the system dynamics simulation when tree-partitioning

approaches are required to handle the corresponding challenges in power system operation, planning, and control.

The clustering (i.e., partitioning, islanding, and splitting) of electric power systems is a concept that emerges especially frequently in control and protection technologies (Park and Kim, 2006; Sarajpoor et al., 2021). In power systems, clustering methods are used in various ways and for various purposes. Chai et al. proposed a double-layer voltage control strategy to solve the voltage violation problem in a distribution system. The proposed method was aimed at minimizing the PV active power curtailment and network power loss of each cluster. In this case, the process of dividing clusters was based on a clustering method that operated according to electrical distance, thus facilitating rapid optimization. In general, following the high penetration of PV power, numerous approaches have been proposed based on the corresponding voltage (for example, to avoid voltage violations). In this context, decentralized or distributed control based on power network clustering offers an advantage over centralized control (Cao et al., 2021). The clustering method can be applied not only to system control but also to event detection (Ma et al., 2021). Nevertheless, in the case of a large existing system, it is difficult to search for events owing to the large quantities of measured data. A graph-theory-based network partitioning algorithm has been proposed to address this problem and accelerate the system response.

The clustering algorithms used in several of the papers cited herein can be classified into different categories (Si et al., 2021). Among these algorithms, hierarchical spectral clustering is an important clustering method and is widely used for the partitioning of electric power systems. Owing to its strong theoretical basis (von Luxburg, 2007; Lee et al., 2014), this method has been used in various power system studies. This spectral clustering technique can be applied using the electrical parameters (e.g., the topology, admittance, and power flow) of a power system alongside hierarchical clustering, where the preferred number of clusters is considered the input. Consequently, hierarchical clustering can be obtained for the connection strength determined by the chosen electrical weighting (Sánchez-García et al., 2014). This method has been studied from various perspectives, such as to improve the calculation efficiency and partition quality through orthogonal conversion (Tyuryukanov et al., 2018), to island power systems according to the minimum active flow disruptions (Amini et al., 2020), and to prevent cascading failures through tree partitioning (Bialek and Vahidinasab, 2022). As described above, numerous studies have focused on clustering; these studies and their findings are summarized and compared in **Table 1**.

However, methods for clustering in power transmission systems that incorporate variable power sources (e.g., renewable generation) have not been fully discussed. This is because when applying power flow-based clustering, the results can differ in response to variations in power generation. Clustering according to topology or admittance can be considered; however, this leads to further issues that cannot be addressed (e.g., those concerning a branch list with an overload risk, maximum overload rates, and overload probabilities). Accordingly, a probabilistic flow analysis method has been proposed to address these issues (Zhu et al., 2020; Wang et al., 2021; Lin et al., 2022). It is necessary to discuss clustering based on a consideration of a probabilistic interpretation of renewable power generation sources (e.g., PV sources). Therefore, in this paper, we propose a probabilistic clustering methodology for power systems operating with a high penetration of VRE sources.

The main contributions of this paper are as follows:

- A probabilistic spectral clustering methodology based on the Monte Carlo method: This method can be applied to power systems by considering the characteristics of VRE

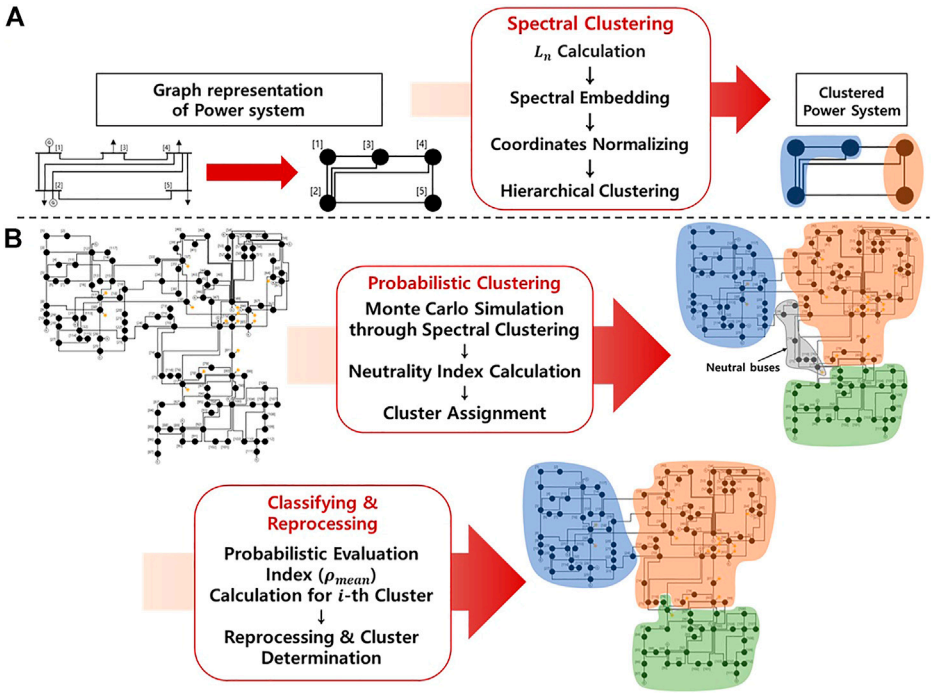

FIGURE 1 | Clustering process. **(A)** Hierarchical spectral clustering method. **(B)** Probabilistic spectral clustering and classification of neutral buses.

sources whilst applying the hierarchical spectral clustering method in the existing power system network.

- The classification and reprocessing of neutral buses: Neutral buses are generated as a result of the proposed probabilistic clustering method and the change in power flow. By reprocessing such buses using neutrality and probabilistic evaluation index, optimal clustering results can be derived.
- The definition of new evaluation indices for probabilistic clustering: The probabilistic expansion index is discussed anew, allowing the expansion to be used as a clustering evaluation index for application under probabilistic conditions. As a result, a more appropriate clustering can be derived.

The remainder of this paper is organized as follows: In **Section 2**, the essential preliminaries are introduced, and the spectral clustering method and related contents are summarized. **Section 3** describes the proposed probabilistic spectral clustering method, PV modeling, and neutral bus classification. In **Section 4**, the results of the method applied to an IEEE 118-bus test system are verified. Finally, the discussion and conclusion are presented in **Section 5** and **Section 6**.

2 HIERARCHICAL SPECTRAL CLUSTERING OF POWER SYSTEM

A power system can be partitioned via hierarchical spectral clustering, as shown in **Figure 1A**; the proposed probabilistic spectral clustering is shown in **Figure 1B**. The spectral clustering algorithm consists of four steps, as illustrated in **Figure 1**.

2.1 Graph Representation of Network

2.1.1 Terminology

For graph-theory-based network partitioning, power systems with N buses are represented as a graph G = (V, E) with a vertex set V and edge set E. The buses and transmission lines (or transformers) in power grids can be denoted as vertices (nodes) and edges (links), as follows:

$$v_i \in V, \qquad i = 1, 2, \ldots, N, \tag{1}$$

$$e_{ij} \in E \subset V \times V, \qquad i, j = 1, 2, \ldots, N. \tag{2}$$

This graph is only considered a simple graph (i.e., no loops or multiple edges are allowed). Multiple edges are replaced by equivalent single edges. All graphs are undirected.

2.1.2 Edge Weights Reflecting Power System

The topological structure of the graph does not include the electrical information of the power system. Therefore, the edge weights should be used. Edge weight is a function w: $E \rightarrow \mathbb{R}$ such that the weight is calculated as follows:

1) $w(e_{ij}) = w(e_{ji})$ for all e_{ij},
2) $w(e_{ij}) = 0$ if $e_{ij} \notin E$,
3) $w(e_{ii}) = \sum_{j=1, j \neq i}^{N} w(e_{ij})$.

In this study, we adopt the notation that $w(e_{ij}) = w_{ij}$ if $i \neq j$, and $d_i = w(e_{ii})$ for the weighted vertex degree. In a purely

TABLE 2 | Characteristics of the photovoltaic (PV) module.

PV module characteristic	Value
Current at maximum power point, I_{MPP} (A)	7.76
Voltage at maximum power point, V_{MPP} (V)	28.36
Short circuit current, I_{SC} (A)	8.38
Open circuit voltage, V_{oc} (V)	36.96
Current temperature coefficients, K_i (A/°C)	0.00545
Voltage temperature coefficients, K_v (V/°C)	0.1278

topological structure, all edges have unit weights. To represent power grids, we can use the values of the power flows as edge weights. When P_{ij} denotes an active power flow from buses i to j, the corresponding edge weight is defined as follows:

$$w_{ij} = \frac{|P_{ij}| + |P_{ji}|}{2}. \tag{3}$$

The power-flow-based weight depends on the operating point and denotes the importance of a branch. Thus, a branch with a small flow is more likely to be removed. In contrast, a branch with a large flow exhibits a strong connection between each vertex and is more likely to be grouped together with other elements.

2.1.3 Graph Laplacian Matrix

Laplacian matrices are used in the spectral clustering method for efficient graph partitioning. The method uses the eigenvector and eigenvalues of two types of Laplacian matrices, which are related to the undirected weighted simple graph G = (V, E, w).

The Laplacian matrix L of G is an $N \times N$ matrix, where N is the number of vertices.

$$[L]_{ij} = \begin{cases} d_i, & if\ i = j \\ -w_{ij}, & if\ i \neq j\ and\ e_{ij} \in E \\ 0, & otherwise. \end{cases} \tag{4}$$

The normalized Laplacian matrix is calculated using a diagonal matrix D with nonzero elements d_i.

$$L_{nor} = D^{-1/2} L D^{1/2}, \tag{5}$$

$$[L_n]_{ij} = [L_n]_{ij} = \begin{cases} 1, & if\ i = j \\ \frac{-w_{ij}}{\sqrt{d_i}\sqrt{d_j}}, & if\ i \neq j\ and\ e_{ij} \in E \\ 0, & otherwise. \end{cases} \tag{6}$$

The normalized Laplacian matrix is scale-independent and more suitable for clustering.

The eigenvalues of the Laplacian matrix are non-negative real numbers, and the number of zero eigenvalues is equal to the number of connected components in the graph. For the normalized Laplacian matrix of a connected graph with N vertices, the eigenvalues can be written as follows (von Luxburg et al., 2008):

$$0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k \leq \ldots \leq \lambda_N \leq 2, \tag{7}$$

$$\lambda_2 > 0. \tag{8}$$

By taking the k smallest eigenvalues together with their respective eigenvectors $(\nu_1, \nu_2, \ldots, \nu_k, \ldots, \nu_N)$ of the normalized Laplacian matrix, we can partition the power system using a spectral clustering method. This method is described in the next section.

2.2 Spectral Clustering

Clustering refers to the classification of the vertices in a graph into several groups (clusters) such that vertices in the same cluster are highly interconnected but are weakly connected to vertices in other clusters. Among the clustering methods, spectral clustering uses normalized Laplacian eigenvalues and eigenvectors. The concept applies the first k eigenvectors corresponding to the smallest k eigenvalues (called spectral k-embedding) and identifies the geometric coordinates that match the N vertices. These coordinates form the N rows of the $N \times k$ matrix, consisting of k eigenvectors. The resulting points are clustered using a spectral clustering algorithm, as discussed below.

2.2.1 Laplacian Matrix Calculation

The Laplacian and normalized Laplacian matrix are calculated for graph G using **Eqs (4)–(6)**, as described in the previous section. At this time, the edge weight is calculated as a power flow using **Eq. (3)**. The latter of the two calculated matrices is used in this study. The eigenvectors of this matrix provide the coordinates representing each bus in the space $\mathbb{R}^k$. The constant k can be determined using the spectral embedding process described in the next section.

2.2.2 Spectral Embedding

Spectral embedding is an important process in spectral clustering. This process employs the first k eigenvectors of the Laplacian matrix, that is, it reduces the dimensions of the matrix from the $N \times N$ Laplacian graph to the $N \times k$ ($k \ll N$) matrix X, which consists of the chosen eigenvectors. In other words, the dimensions to be analyzed are reduced from $\mathbb{R}^N$ to $\mathbb{R}^k$. This makes it possible to increase the quality of the resulting clusters whilst reducing the computation time.

The criterion for determining the embedding space k is based on the eigengaps (i.e., the differences between two sequential eigenvalues). However, in general, we can obtain a better k -partition with smaller eigenvalues (Sánchez-García et al., 2014). Hence, we use a relative eigengap, as follows:

$$\gamma_k = \frac{\lambda_{k+1} - \lambda_k}{\lambda_k} \qquad (k \geq 2). \tag{9}$$

When the value of the eigengap is large, we obtain a better network partition with at least k clusters. Therefore, in the following simulation, the value of k that maximizes the eigengap value is set as the embedding space.

2.2.3 Coordinate Normalization

After spectral embedding, the coordinates of the N vertices are represented in space $\mathbb{R}^k$. However, the magnitudes of coordinates differ; hence, the results from the clustering may be distorted.

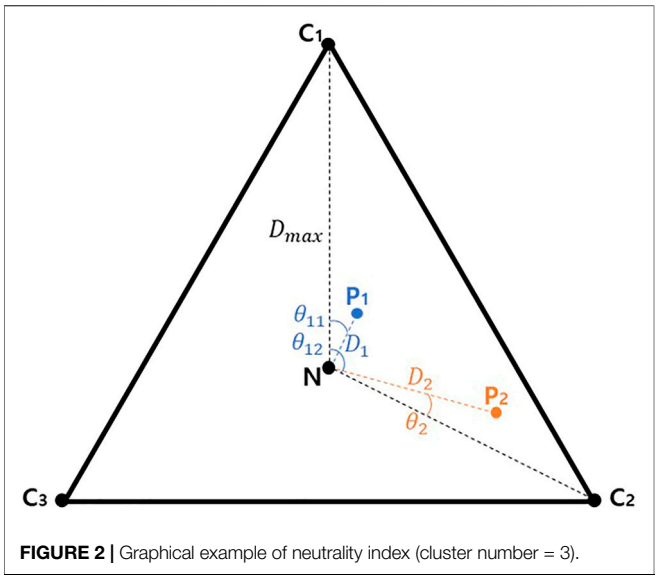

FIGURE 2 | Graphical example of neutrality index (cluster number = 3).

Therefore, spectral clustering is performed once each coordinate value has been normalized to norm 1. This process involves projecting each coordinate into a hypersphere, i.e., a (k -1)-sphere (depending on the embedding space k). The value of each coordinate can be normalized as follows:

$$u_i = \frac{x_i}{\|x_i\|} \quad (1 \le i \le N). \tag{10}$$

2.2.4 Hierarchical Clustering

Through the clustering process described above, we express each bus as a coordinate in space $\mathbb{R}^k$, and we perform clustering by classifying the coordinates. Several methods are available for grouping each bus, including k-means, k-medoid, and hierarchical clustering (Sarajpoor et al., 2021). The k-means and k-medoid methods are frequently used for grouping, but they have several disadvantages. First, the clusters must be determined before clustering. Moreover, the connections between buses in the graph are neglected. Thus, the hierarchical clustering method is applied in this study.

Hierarchical clustering offers several advantages over other clustering methods. It identifies the hierarchy of clusters, which we represent as dendrograms. A dendrogram is a tree diagram that visualizes the hierarchical spectral structure of a power system. This method employs a bottom-up process applied for power system clustering. In N buses, the two most similar buses are merged into a cluster. Here, similarity refers to the distance between buses or clusters. This distance indicates the Euclidian distance between buses i and j.

$$similarity_{ij} = \left\|u_i - u_j\right\| \quad \left(u_i, u_j \in \mathbb{R}^k\right). \tag{11}$$

After the buses are chosen, a new graph with N-1 buses is formed. In the new graph, the two closest buses are merged into a new cluster again, and this process is repeated until the desired number of clusters is obtained. This procedure helps grasp the hierarchical structure immediately, and the number of clusters can be adjusted without additional calculations.

2.3 Clustering Evaluation

Through the hierarchical spectral clustering described thus far, the desired number of clusters can be obtained from graph G. The purpose of clustering is to identify a cluster in which the buses in the cluster are strongly connected to each other and weakly connected to buses in other clusters. To evaluate the quality of a cluster, two quantities are defined to evaluate the connectivity described above. First, boundary (∂) is defined as the sum of the edge weights of buses within a specific cluster C and those of the buses within other clusters. It can be used to evaluate the connectivity between different clusters.

$$\partial(C) = \sum_{i \in C,\ j \notin C} w_{ij}. \tag{12}$$

Second, the volume (vol) of cluster C is defined as the sum of the degrees of cluster C. It can be used to evaluate the connectivity of the buses within a specific cluster:

$$vol(C) = \sum_{i \in C} d_i. \tag{13}$$

For the abovementioned purposes, the quality evaluation metric for a cluster can be defined as the value obtained by dividing the boundary by the volume. It is denoted as the expansion (ϕ) (Sarajpoor et al., 2021).

$$\phi(C) = \frac{\partial(C)}{vol(C)}. \tag{14}$$

The smaller the value of the expansion for the network partition, the better the clustering (i.e., strong connections between buses included in a specific cluster, and weak connections between buses outside thereof). When partitioning the power system into k clusters, we use the maximum expansion value to evaluate the quality of the clustering method (Sánchez-García et al., 2014):

$$\phi_{max}(C_1, \ldots, C_k) = \max_{1 \le i \le k} \phi(C_i). \tag{15}$$

It is possible to evaluate whether clustering has been well executed by using a normalized cut (i.e., the arithmetic mean of all clusters' expansions) (von Luxburg, 2007). This method is widely used in the evaluation of graph clustering because the average quality of all clusters can be evaluated.

$$\phi_{mean}(C_1, \ldots, C_k) = \frac{1}{k} \sum_{i=1}^{k} \phi(C_i). \tag{16}$$

In this study, in view of the expansion of renewable energy sources in power systems, a metaheuristic approach is applied for probabilistic interpretation. To this end, numerous clustering results must be analyzed for various cases. For this purpose, the above two objective function equations are used to efficiently optimize and evaluate the results. In contrast, several more complex objective functions were required for the method proposed by Cotilla-Sanchez et al. (2013).

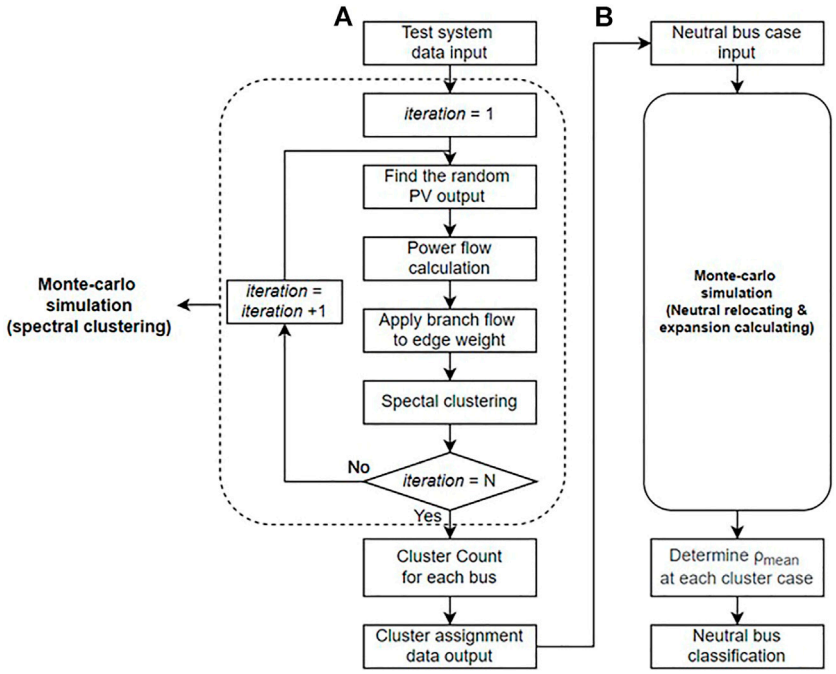

FIGURE 3 | Proposed methodology. **(A)** Probabilistic spectral clustering algorithm. **(B)** Neutral bus classification.

3 PROBABILISTIC SPECTRAL CLUSTERING METHODOLOGY

In this section, we propose a probabilistic spectral clustering method applicable to power systems containing variable sources (e.g., renewable energy sources). In addition, we explain the process of optimizing the neutral buses that may arise in the process of probabilistic analysis by using an evaluation index (i.e., the expansion described in the previous section). To this end, a model is required for the PV output power and its corresponding variability.

3.1 Modeling of Variable Renewable Energy Source: Photovoltaic (PV)

In this study, assuming a situation in which PV sources are expanded in a power system, we repeatedly perform the clustering process for a power flow with an intermittent PV output. The output power of the PV is affected by the random phenomenon of solar irradiance, which changes the power flows of the branches in the power system. Generally, it is more efficient to model irradiance using a beta probability density function than other probability density functions (Teng et al., 2013). Accordingly, the solar irradiance can be expressed as follows:

$$f_{beta}(s) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} s^{\alpha-1}(1-s)^{\beta-1} & (0 \le s \le 1, \quad \alpha, \beta \ge 0) \\ 0 & otherwise \end{cases}, \tag{17}$$

$$\beta = (1-\mu)\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right), \tag{18}$$

$$\alpha = \frac{\mu\beta}{1-\mu}. \tag{19}$$

In the above, $f_{beta}(s)$ denotes the beta distribution function for solar irradiance, where s is a random variable characterizing the solar irradiance (kW/m^2); Γ is a gamma function comprising the beta distribution function; and α and β are the parameters of the beta distribution function, where both must be positive. μ and σ denote the mean and standard deviation of s, respectively, and these values are used to calculate the parameters. The values of the mean and standard deviation of s for a specified time period are 0.657 kW/m^2 and 0.284 kW/m^2, respectively (Hung et al., 2014).

The output power of the PV module depends on the solar irradiance, ambient temperature, and parameters of the PV module. When the PV module operates at the maximum power point and at a solar irradiance s, the output power can be calculated as a function of s as follows (Sehsalar et al., 2019):

$$P_{PV}(s) = N \times FF \times V(s) \times I(s), \tag{20}$$

$$FF = \frac{V_{MPP} I_{MPP}}{V_{oc} I_{SC}}, \tag{21}$$

$$V(s) = V_{oc} - K_V \times T_C, \tag{22}$$

$$T_C = T_a + s \times \frac{T_n - 20\ (℃)}{0.8}, \tag{23}$$

$$I(s) = s \times (I_{SC} + K_i \times (T_C - 25\ (℃)). \tag{24}$$

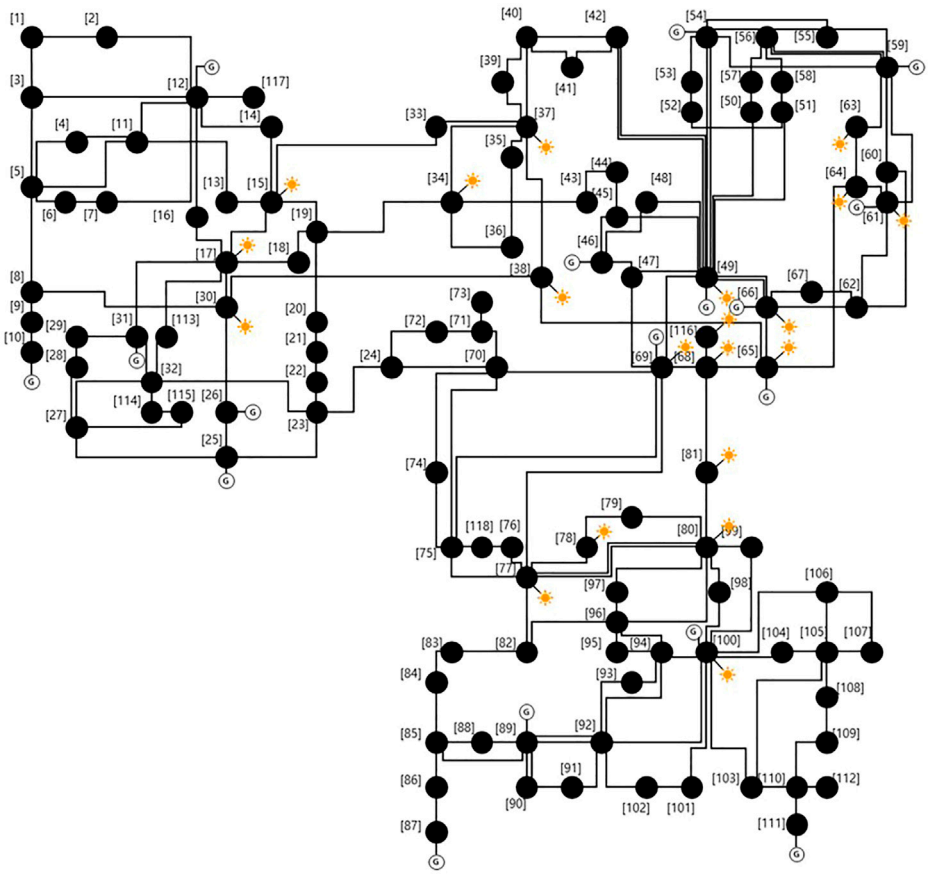

FIGURE 4 | IEEE 118-bus system with photovoltaic (PV) integration.

Here, N denotes the number of PV modules, and FF is the fill factor obtained from **Eq. (22)**; V_{MPP} and I_{MPP} are the voltage and current at the maximum power point in V and A, respectively; V_{oc} and I_{SC} are the open-circuit voltage and short-circuit current, respectively; T_C, T_a, and T_n are the cell, ambient, and nominal operating temperatures of the PV cell, respectively (in °C); and K_V and K_i are the voltage and current temperature coefficients, respectively (in V/°C and A/°C, respectively). The output power of the PV system installed on the buses in the power system can be determined using **Eq. (20)**. The parameters of the PV module in **Table 2** also reflect the values reported in the study by Hung et al. (2014).

3.2 Probabilistic Optimal Power Flow Algorithm

The PV output varies depending on the changes in the solar irradiance over time; accordingly, the value of the power flow in the power system also varies. Owing to this randomness, the results from the power-flow-based spectral clustering vary continuously. The power generation of the existing turn-on generators should be readjusted to match the PV output. The total power generation of conventional generators is adjusted via the total load, PV generation, and power loss parameters, as follows:

$$\sum_{i=1}^{n_g} P_{Gi} = \sum_{j=1}^{n_l} P_{Lj} - \sum_{k=1}^{n_{PV}} P_{PVk} + P_{loss}. \tag{25}$$

In the above, P_{Gi} denotes the active power generation at the generator bus i; P_{Lj} is the load at the bus j; P_{PVk} denotes the PV generation at bus k; n_g, n_l, and n_{PV} denote the total number of generator buses, load buses, and buses connected to PV, respectively; and P_{loss} is the power loss in the network.

In this study, the problem regarding the re-dispatching of generators is solved using an optimal power flow (OPF) calculation, and the objective function is the total fuel cost, as represented by the generator output active power (Chayakulkheeree, 2014; Shaheen et al., 2019).

$$\min_{P_{Gi}} FC = \sum_{i=1}^{n_g} cost\,(P_{Gi}). \tag{26}$$

In the above, FC denotes the total cost of the conventional generator connected to bus i, n_g is the total number of generator buses and P_{Gi} denotes the active power generated at generator bus i.

The constraints of the OPF problem are the power-balance equation and variable limits.

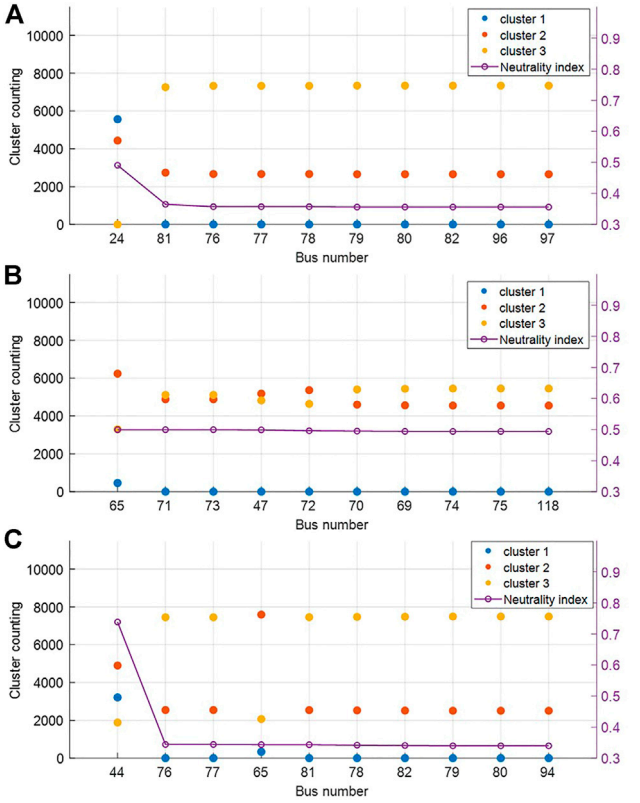

FIGURE 5 | Simulation result at **(A)** 20%, **(B)** 50%, **(C)** 80% penetration level.

$$P_{inj,k} - \sum_{l=1}^{N} V_k V_l [G_{kl} \cos(\delta_l - \delta_k) + B_{kl} \sin(\delta_l - \delta_k)] = 0, \quad (27)$$

$$Q_{inj,k} - \sum_{l=1}^{N} V_k V_l [G_{kl} \sin(\delta_l - \delta_k) - B_{kl} \cos(\delta_l - \delta_k)] = 0. \quad (28)$$

Here, $P_{inj,k}$ is the total active power injected into bus k, and $Q_{inj,k}$ is the total reactive power injected into bus k. G_{kl} and B_{kl} are the conductance and susceptance of the admittance Y_{kl} (an element of the admittance matrix), respectively; δ_l and δ_k are the voltage angles at buses k and l, respectively; and N is the total number of buses. Thus,

$$P_{Gi,min} \le P_{Gi} \le P_{Gi,max}, \qquad i = 1, 2, \ldots, n_g, \quad (29)$$

$$Q_{Gi,min} \le Q_{Gi} \le Q_{Gi,max}, \qquad i = 1, 2, \ldots, n_g, \quad (30)$$

$$V_{i,min} \le V_i \le V_{i,max}, \qquad i = 1, 2, \ldots, n, \quad (31)$$

$$V_k V_l [G_{kl} \cos(\delta_l - \delta_k) + B_{kl} \sin(\delta_l - \delta_k)] \le flow_{\lim,kl}, \quad k, l = 1, 2, \ldots, n. \quad (32)$$

In the above equation, $flow_{lim,kl}$ is the branch flow limit of the line that connects buses k and l.

The clustering considering the random variable should be performed by aggregating the repeated results obtained from the power-flow-based clustering affected by the random PV output. To this end, we propose a probabilistic spectral clustering algorithm. This algorithm repeats the spectral clustering process described in the previous chapter (**Figure 1**) by reflecting the power flow determined from the power flow calculations considering the variable PV output. This algorithm is summarized as follows.

1) Employ test power system data for the probabilistic spectral clustering algorithm.
2) Determine the random PV output calculated through **Eq. (20)**. The solar irradiance follows the beta distribution function; hence, the value of the PV output is randomly determined, changing the value of the power flow.
3) Calculate the power flow in the power system whilst considering the PV output as a negative load. The power flow may vary depending on the random PV output at each moment.
4) Apply each branch's flow to the corresponding edge weight. In this study, the weight is based on the active power flow, and the value of the edge weight depends on the PV output.
5) Perform the spectral clustering shown in **Figure 1**. This process involves calculating the Laplacian matrix, spectral embedding, and hierarchical clustering.
6) Repeat Steps 2–5 N times (i.e., the maximum iteration number). The PV output varies at each iteration; hence, the clustering results from the test power system also vary.
7) Count the number of clusters in which each bus is included after the N iterations.
8) Assign buses to each cluster. In this process, non-neutral buses can be easily allocated, though some neutral buses are not. Hence, it is necessary to classify the neutral buses. This is described in the next section.

3.3 Neutral Buses and Classification

3.3.1 Definition of Neutral Buses

Under a probabilistic condition, the clustering result varies according to the variable generation source. This is because the change in PV generation affects the power flow in the power system. As a result, certain buses are grouped into clusters; however, it remains difficult to identify which buses are classified into which clusters. These buses can be referred to as neutral buses. The neutral buses can also be treated as having an almost equal probability of being assigned to each cluster. To define neutrality in this study, the distance between two points is used [i.e., from the most neutral case and from the counting number (in which each bus is grouped in each cluster) represented in the coordinates]. It is possible to identify which cluster has a high grouping probability for each bus by using the angle between the cases grouped into only one cluster and the coordinates of each case. Using these two elements, the neutrality index (NI) at each bus can be defined as follows:

$$D = \left(\sum_{i=1}^{k} |n_i - p_i|^2 \right)^{1/2}, \quad (33)$$

$$NI = \left(\frac{D_{max} - D}{D_{max}} \right)_i. \quad (34)$$

In the above, D denotes the distance between the coordinates P (consisting of the counting number in which

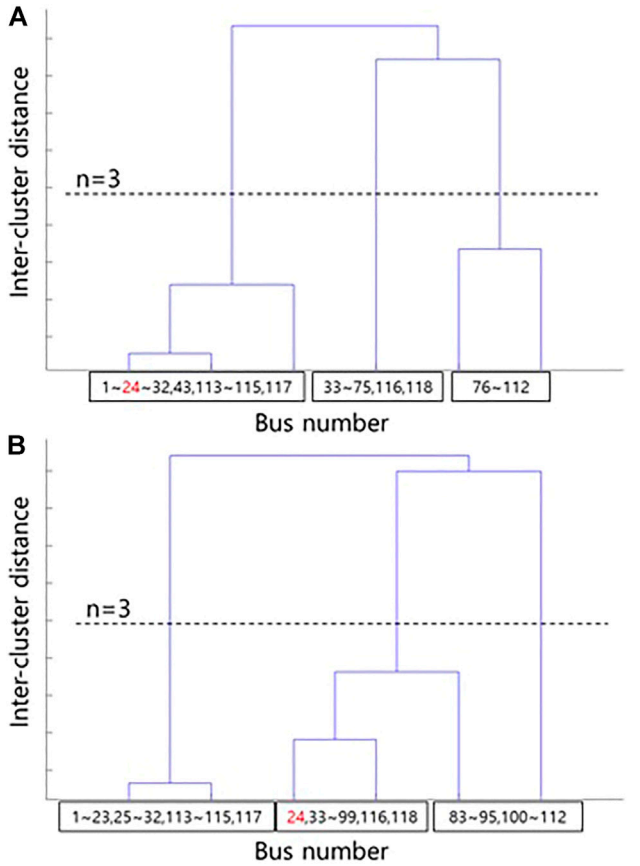

FIGURE 6 | Hierarchical structure at 20% penetration level with respect to PV output. **(A)** Bus 24 belongs to Cluster 1. **(B)** Bus 24 belongs to Cluster 2.

each bus is grouped into each cluster) and the coordinates for the most neutral coordinate N. The distance between the two points N and P is given by the Euclidean distance. If the number of clusters is k for $N(n_1, n_2, \ldots, n_k)$ and $P(p_1, p_2, \ldots, p_k)$, then the distance in k dimensions is defined as shown above.

i indicates the cluster to which each bus belongs with a higher probability. As shown in **Figure 2**, i can be obtained from the angles between the vector NP (formed by the most neutral coordinate N and the coordinates for the bus to be obtained) and the vectors NC_i [formed by N and the least neutral coordinates C_i (i = 1, 2, . . ., k)]. For example, if the cluster number is three, the coordinate of the most neutral case is N (3,333.33, 3,333.33, 3,333.33) when the number of iterations is 1×10^4. Moreover, D_{max} denotes the distance between the least neutral case [C_1 (10,000, 0, 0), C_2 (0, 10,000, 0), or C_3 (0, 0, 10,000)] and the most neutral coordinates. The case of P_1 case has a smaller distance D_1 than that of P_2. Therefore, it can be said that the NI value of P_1 is larger and more neutral. Indicator i can be obtained from the case with the smallest value between the angles θ formed by the two vectors NP_m (m = 1, 2) and NC_n (n = 1, 2, 3). For the coordinate P_1, an angle θ_{11} formed with NC_1 is less than an angle θ_{12} formed with NC_2; thus, the value of i is 1. For this reason, it may be seen that the value of i for P_2 is 2.

For neutral buses with high NI values, a reclassification process and a verification process are required. These processes refer to the evaluation of a neutral bus using an index appropriate for the probabilistic environment, as described in the next section.

3.3.2 Classification and Reprocessing Based on Probabilistic Evaluation Index

To evaluate clustering results (as described in the previous section), we calculate the expansion [**Eq. (14)**] of each cluster and use their maximum (**Eq. (15)**) or average (**Eq. (16)**) expansion values. However, this method can only evaluate the results when clustering the power system in each iteration: it is impossible to evaluate the overall results of the probabilistic clustering (i.e., those reflecting the changing power flow with respect to the variable PV power output). Therefore, a novel method is needed to evaluate the clustering results whilst reflecting random renewable generation during iteration. This method can also be used to determine the clusters into which the neutral buses should be classified.

To evaluate the clustering results from N iterations, the representative index (based on expansion) must be determined in the evaluation. We can consider the maximum and average values of expansion as the index, as described in the previous section (**Eq. 15** and **Eq. 16**). There are N iterations after setting the representative value at each iteration; hence, the probabilistic clustering quality should be evaluated using N data elements. In this study, we use the maximum value amongst the N representatives as a probabilistic clustering evaluation index to conservatively judge the clustering quality when the PV output drastically changes the power current in the system. Therefore, the indices for evaluating the probabilistic clustering results can be defined using the following two equations (with reference to **Eq. 15** and **Eq. 16**):

$$\rho_{max} = \max_{1 \le j \le N} \phi_{max.j}, \tag{35}$$

$$\rho_{mean} = \max_{1 \le j \le N} \phi_{mean.j}. \tag{36}$$

Here, ρ_{max} and ρ_{mean} are the probabilistic clustering evaluation indices based on the expansion maximum (ϕ_{max}) and mean (ϕ_{mean}), respectively. Because the two indices take the maximum value of the expansion in the overall iteration, we can see that even when the PV output has the worst effect on power system clustering, the indices indicate whether the cluster is well established and determine the average cluster quality of all clusters. However, in this study, we did not use the index ρ_{max} to classify neutral buses; instead, we used ρ_{mean}. The evaluation index was defined using the mean expansion value because if the maximum expansion value is used, the expansion value of the worst cluster will take this same maximum value, resulting in a probabilistic index overlap.

The aforementioned index can be used to determine which cluster is the most appropriate for a neutral bus to be grouped into. To this end, the neutral buses are first relocated to each cluster, and the expansion values (which vary depending on the PV output) are aggregated. By calculating the probabilistic

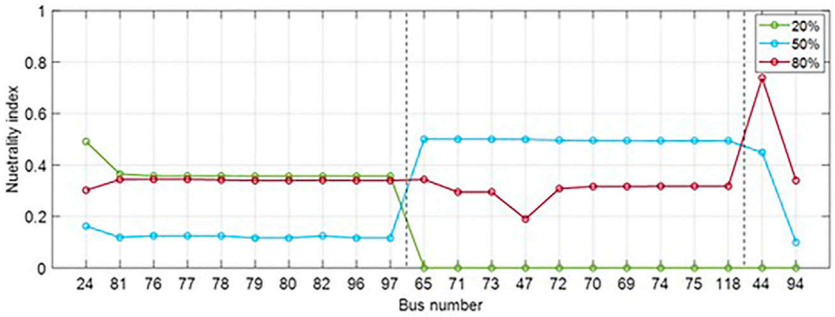

FIGURE 7 | Neutrality index results at each penetration level.

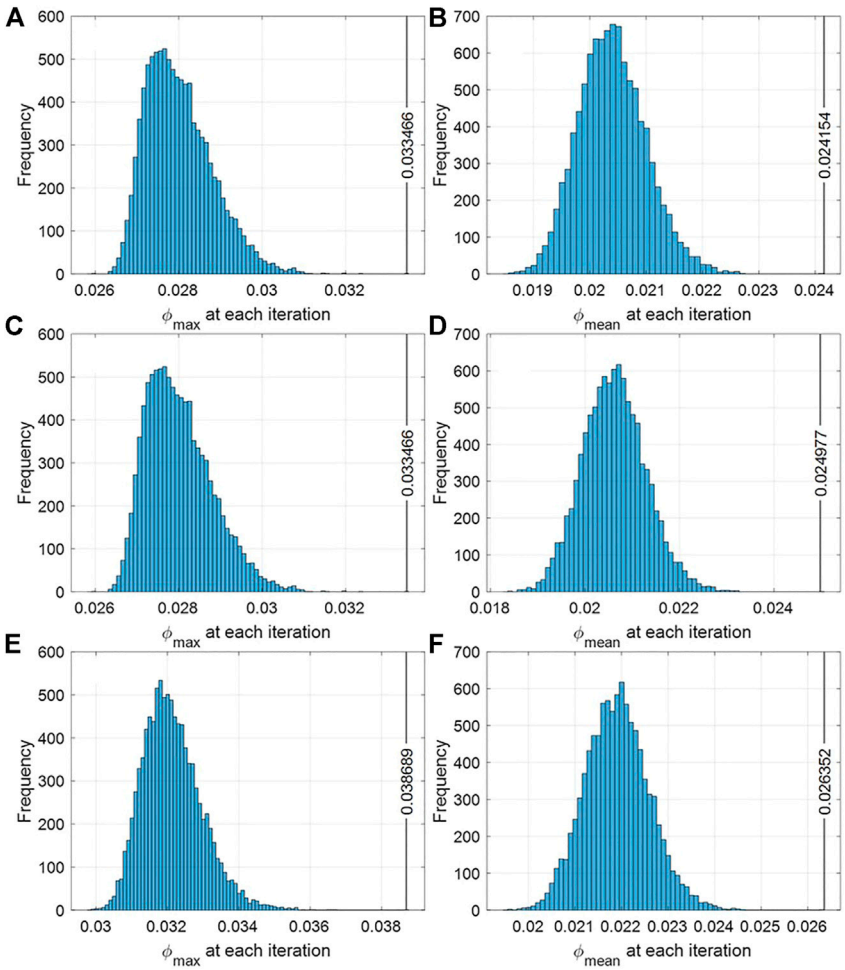

FIGURE 8 | Histogram of probabilistic clustering evaluation indices for each cluster at a 20% penetration level. **(A–C)** ρ_{max} of Clusters 1, 2, and 3, respectively **(D–F)** ρ_{mean} of Clusters 1, 2, and 3, respectively.

clustering evaluation index when placing the neutral buses in each cluster, we can determine in which cluster the value of the index is minimized when placing the neutral buses. At this time, the neutral buses can be classified into their corresponding clusters. That is, the neutral buses can be further classified using the mean index (ρ_{mean}). The flowchart in **Figure 3** summarizes this process and the probabilistic spectral clustering algorithm discussed in the previous section.

TABLE 3 | List of buses connected to PV model.

Bus numbers
15, 17, 30, 34, 37, 38, 49, 61, 63, 64, 65, 66, 68, 69, 77, 78, 80, 81, 100, 116

4 CASE STUDY

As described in this section, the proposed probabilistic spectral clustering algorithm and neutral bus classification were tested using an IEEE 118-bus test system (as shown in **Figure 4**). This system approximates the American Electric Power system and contains 19 generators, 35 synchronous condensers, 177 lines, nine transformers, and 91 loads. The total generator output and load consumption of the test system were 4,374.5 MW and 4,242.0 MW, respectively. To reflect the random outputs of PV generation, the PV models described in the previous section were connected to 20 buses in the power system as shown in **Table 3**. By varying the rating of the PV power to 42.42, 106.05, and 169.68 MW, we tested the probabilistic spectral clustering algorithm for system penetration levels of 20, 50, and 80%, respectively. A penetration level means the ratio of the sum of PV ratings to the total load. Depending on the output of the PV system, the outputs of the generators that were turned on were re-dispatched for the OPF calculations. All numerical calculations (e.g., the power flow calculations) were implemented using MATLAB software (Zimmerman and Murillo-Sanchez, 2020).

We first performed the proposed probabilistic spectral clustering (described in the previous section) based on the power flow Laplacian. Owing to the variable generation source (here, the PV source), the power flow in the test system differed for each iteration. Consequently, the coordinates of each bus differed for each iteration, producing different clustering results. In this simulation, the 118-bus system with PV integration was divided into three clusters. As shown in **Table 4-1**, **Table 4-2**, and **Table 4-3**, we observed the number of times each bus was partitioned into clusters during the 10,000 iterations. Referring to the tables, **Figure 5** shows the graphs for the cluster counting number and the NI [**Eq. (33)**] for the 10 most neutral buses.

Each table summarizes the results from the probabilistic spectral clustering after 10,000 repetitions, and buses with the same counting number were grouped and displayed in ascending order. Therefore, when the count number was divided by 10,000, the probability of the buses belonging to each cluster could be calculated. This shows the number of times each bus is grouped into each cluster in the 118-bus system incorporating a variable PV power output.

From **Table 4-1**, it can be seen that certain buses can be reliably classified into each cluster; however, in the case of Bus 24, the counting numbers grouped into two clusters are similar; hence, it can be considered as a neutral bus. Furthermore, **Figure 6** shows snapshots in which Bus 24 is grouped differently into Clusters 1 and 2 according to the variability of PV; here, the hierarchical structure of the buses (as analyzed by spectral clustering) is shown. The similarity of the buses varies owing to PV variability, which changes the cluster to which Bus 24 is assigned.

Next, the counting numbers in which the buses are clearly classified in **Table 4-1** decreased as shown in **Table 4-2** to **Table 4-3** furthermore, when the penetration level of PV sources increased, reliable clustering became more difficult. In particular, in the case of a 20% penetration level, Buses 44 and 65 were reliably clustered; however, when the penetration level increased to 50 and 80%, the buses become neutral and clustering them presented ambiguity.

The simulation results at each penetration level suggest that different neutral buses appear in each case. The most neutral buses for each case were obtained from the maximum NI. Accordingly, the NIs grouped into the clusters for the buses described in **Figure 5** at each penetration level are summarized in **Figure 7**. Among the three clusters, the values of the maximum cluster probability and NIs for most neutral buses at each penetration level are listed in **Table 5**.

As shown in **Figure 7**, the NIs for neutral buses varied according to the penetration level. In particular, when the penetration level increased, the NI value of the most neutral bus likewise increased. The result of the clustering was not ensured, owing to the variable power generation (e.g., with PV sources). In addition, in **Table 5**, by observing the number of buses exhibiting the maximum probability for each penetration level, we see that buses were most neutral at penetration levels of 20 and 80%. At a penetration level of 50%, the probability of the most neutral bus (Bus 65, which had the largest NI value) was not the smallest value. This is because the counting number of Bus 65 was grouped into Cluster 1, and the value of NI also increased.

At a penetration level of 20% or 80%, the probability of each bus being grouped into a specific cluster was less than or approximately equal to 50%. Furthermore, the higher the penetration level, the lower the probability. A low clustering

TABLE 4-1 | Simulation results at 20% penetration level.

Bus number	Cluster 1	Cluster 2	Cluster 3	Bus number	Cluster 1	Cluster 2	Cluster 3
1–23	10,000	0	0	83–95	0	0	10,000
24	5,564	4,436	0	96–99	0	2,657	7,343
34–75	0	10,000	0	100–112	0	0	10,000
76–78	0	2,668	7,332	113–115	10,000	0	0
79–80	0	2,657	7,343	116	0	10,000	0
81	0	2,739	7,261	117	10,000	0	0
82	0	2,657	7,343	118	0	10,000	0

TABLE 4-2 | Simulation results at 50% penetration level.

Bus number	Cluster 1	Cluster 2	Cluster 3	Bus number	Cluster 1	Cluster 2	Cluster 3
1–13	10,000	0	0	66–67	0	6,338	3,662
14–18	9,997	3	0	68	0	6,245	3,755
19	9,992	8	0	69	0	4,562	5,438
20–23	9,997	3	0	70	0	4,599	5,401
24	1,117	8,883	0	71	0	4,884	5,116
25–32	9,997	3	0	72	1	5,363	4,636
33	1,665	7,542	793	73	0	4,884	5,116
34	1,665	7,548	787	74–75	0	4,550	5,450
35	1,665	7,533	802	76–77	0	849	9,151
36	1,665	7,548	787	78	0	846	9,154
37–40	1,665	7,533	802	79–80	0	788	9,212
41	1,502	7,618	880	81	0	807	9,193
42	0	6,338	3,662	82	0	848	9,152
43	1,664	7,547	789	83	0	788	9,212
44	0	6,340	3,660	84–94	0	671	9,329
45	0	6,338	3,662	95–99	0	788	9,212
46	0	6,023	3,977	100–112	0	671	9,329
47	0	5,181	4,819	113–115	9,997	3	0
48–63	0	6,338	3,662	116	0	6,245	3,755
64	17	6,328	3,655	117	10,000	0	0
65	461	6,237	3,302	118	0	4,550	5,450

TABLE 4-3 | Simulation results at 80% penetration level.

Bus number	Cluster 1	Cluster 2	Cluster 3	Bus number	Cluster 1	Cluster 2	Cluster 3
1–23	10,000	0	0	73	2	7,873	2,125
24	2,185	7,815	0	74–75	2	7,683	2,315
25–32	10,000	0	0	76–77	2	2,546	7,452
33–40	9,928	70	2	78	1	2,527	7,472
41	9,818	178	4	79e80	1	2,509	7,490
42	181	7,733	2086	81	1	2,541	7,458
43	9,927	71	2	82	2	2,517	7,481
44	3,213	4,901	1886	83–84	0	2079	7,921
45	181	7,733	2086	85–92	0	2078	7,922
46	89	8,593	1,318	93	0	2,104	7,896
47	5	8,682	1,313	94–99	0	2078	7,922
48–64	181	7,733	2086	100	0	2,508	7,492
65	337	7,594	2069	101	0	2,500	7,500
66–67	181	7,733	2086	102	0	2078	7,922
68	11	7,830	2,159	103–112	0	2,508	7,492
69	2	7,691	2,307	113–115	10,000	0	0
70	2	7,692	2,306	116	11	7,830	2,159
71	2	7,873	2,125	117	10,000	0	0
72	123	7,809	2068	118	2	7,682	2,316

probability does not guarantee that the buses grouped into a specific cluster represent the optimal clustering results. Therefore, when neutral buses are assigned to three clusters, it is necessary to determine which cluster is most appropriate for grouping, by considering the variable PV power output. To this end, the expansion-based evaluation process described in **Figure 3** is required. In this example, the classification process was applied to Buses 24, 65, and 44, and the index ρ_{mean} was calculated. The smaller the value of each index, the smaller the value of the expansion, indicating a more optimal clustering result. In **Figure 8** (below), histograms of the values ϕ_{max} and ϕ_{mean} (as calculated using the Monte Carlo procedure) are shown. The maximum values for each index (i.e., the values of ρ_{max} and ρ_{mean}) are also shown in the figure.

Neutral bus classification is the process of obtaining the expansion value for each cluster when placing neutral buses into Clusters 1–3 (for each penetration level) and repeating this process 10,000 times by considering the random output of the PV power. As a result, when each neutral bus was assigned to Clusters 1 and 2, the evaluation index value was minimized; that is, optimal clustering was achieved. **Figure 8A** and **Figure 8B** show the same histogram for the penetration

TABLE 5 | Most neutral buses in each penetration level case.

Penetration level (%)	Neutral buses	Probability	Neutrality index
20	24	55.64% (cluster 1)	0.4905_1
50	65	62.37% (cluster 2)	0.4998_2
80	44	49.01% (cluster 2)	0.7383_2

TABLE 6 | Probabilistic clustering evaluation index in each penetration level case. The minimum value for the evaluation index ρmean. Since the cluster with the minimum ρmean is the optimal cluster, we emphasized this with bold fonts.

Penetration level	20 (%)		50 (%)		80%	
Evaluation index	ρ_{max}	ρ_{mean}	ρ_{max}	ρ_{mean}	ρ_{max}	ρ_{mean}
Cluster 1	3.3466	**2.4154%**	10.8802	7.1623%	7.6356%	5.1725%
Cluster 2	3.3466	2.4977%	4.6483	**3.2697%**	7.4233%	**5.1259%**
Cluster 3	3.8689	2.6325%	10.8802	7.2417%	7.6356%	5.2654%

level of 20% (shown in **Figure 7**). It can be seen that when ϕ_{max} used, the distribution of the expansion values is completely consistent, making it impossible to classify neutral buses using ρ_{max}. In addition, because the shapes and maximum values of ρ_{mean} in **Figures 8D, E**, and **Figure 8F** differ from each other, it is necessary to check the value of ρ_{mean} and classify the neutral buses.

In the analysis results, the optimal cluster for each neutral bus was found to be the cluster with the highest probability in **Table 5**. The optimal cluster can be obtained using this value because the value of indicator i matches the cluster with the highest probability.

5 DISCUSSION

In this simulation, when the penetration level increased, the probability that most buses were grouped into a specific cluster also decreased. Thus, as the penetration level of PV power increased, the clustering results changed more frequently, and the number of buses that were difficult to classify increased. The NI value for each bus also increased. Therefore, the most neutral buses can be found using the NI values, and such buses should be classified.

The most neutral buses for each penetration level can be organized as shown in **Table 5**. As a result, at penetration levels of 20 and 80%, the probability of an optimal cluster is low (~50%), and the NI is high. This is because differences between the probabilistic evaluation index values for each cluster in **Table 6** are very small; thus, each clustering result has a similar effect. For this reason, to obtain the most optimal case among similar clustering results, it is necessary to group the clusters with the highest clustering probability among the neutral buses, as well as to verify the results using the probabilistic clustering evaluation index.

In the case of a penetration level of 50%, the NI value was high; however, the maximum probability was also high. When applying the neutral bus classification process in this case, the differences in the values of the probability evaluation index between Cluster 2 and other clusters were found to be large. Therefore, classification should be first performed on the most neutral buses (with the smallest maximum probability), before proceeding to the next-most neutral buses. Thus. buses will be grouped into clusters with the highest probability, similar to the results mentioned above.

6 CONCLUSION

This paper introduces a novel method for performing spectral clustering in power systems with variable sources (e.g., renewable energy sources). Because the output of a renewable energy source is variable, it is necessary to consider variable generation at every moment when clustering. To this end, we propose a probabilistic approach based on a widely adopted graph-theory-based spectral clustering method. The conventional spectral clustering method is limited in that it can only be used in power systems for which the power generation and load can be determined. To overcome this problem and reflect the effects of variable power sources, we developed the probabilistic spectral clustering method to group buses into clusters with the highest grouping probability. The results from the probabilistic clustering can be further optimized by classifying neutral buses. The proposed method was applied to an IEEE 118-bus system with PV integration to confirm the clustering results, which varied depending on the PV penetration level. In addition, the neutral buses for each case generated by the probabilistic spectral clustering method were classified to calculate the probabilistic evaluation index for each cluster and to verify which cluster yielded the optimal clustering.

This method can be applied to power system clustering under conditions of variable renewable energy sources, and it is expected to be implemented in power system planning and operation in the future for net-zero establishment. The probabilistic clustering methodology applied in this study synthesizes the expansion value, which depends on the variability of VRE, via the Monte Carlo procedure and

identifies the optimal clustering point. This probabilistic analysis method can help determine an optimal operation plan in environments where the penetration of renewable energy is high. In particular, from the perspective of system planning and operation, improvements in system reliability and stability can be expected in environments with high renewable energy penetration. This can be achieved through the installation and control of high-voltage direct current across the interfaces between clusters determined using the probabilistic clustering methodology.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

JK put forward the main research points and mathematical analysis and wrote the manuscript. SK, JL, and SH conceived the idea and contributed to data interpretation. MY led the experiments. MY and GJ gave guidance on the study and wrote the manuscript.

FUNDING

This research was supported by the Basic Research Program through the National Research Foundation of Korea (NRF), funded by the MSIT (No. 2020R1A4A1019405), as well as a Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korean Government (MOTIE) (No. 20191210301890).

REFERENCES

Amini, M., Samet, H., Seifi, A. R., Al-Dhaifallah, M., and Ali, Z. M. (2020). An Effective Multi-Solution Approach for Power System Islanding. *IEEE Access* 8, 93200–93210. doi:10.1109/ACCESS.2020.2995085

Bialek, J. W., and Vahidinasab, V. (2022). Tree-partitioning as an Emergency Measure to Contain Cascading Line Failures. *IEEE Trans. Power Syst.* 37, 467–475. doi:10.1109/TPWRS.2021.3087601

Cao, D., Zhao, J., Hu, W., Ding, F., Huang, Q., Chen, Z., et al. (2021). Data-driven Multi-Agent Deep Reinforcement Learning for Distribution System Decentralized Voltage Control with High Penetration of PVs. *IEEE Trans. Smart Grid* 12, 4137–4150. doi:10.1109/TSG.2021.3072251

Cauz, M., Bloch, L., Rod, C., Perret, L., Ballif, C., and Wyrsch, N. (2020). Benefits of a Diversified Energy Mix for Islanded Systems. *Front. Energy Res.* 8, 147. doi:10.3389/fenrg.2020.00147

Chai, Y., Guo, L., Wang, C., Zhao, Z., Du, X., and Pan, J. (2018). Network Partition and Voltage Coordination Control for Distribution Networks with High Penetration of Distributed PV Units. *IEEE Trans. Power Syst.* 33, 3396–3407. doi:10.1109/TPWRS.2018.2813400

Chayakulkheeree, K. (2014). "Probabilistic Optimal Power Flow: An Alternative Solution for Emerging High Uncertain Power Systems," in 2014 International Electrical Engineering Congress (iEECON), Pattaya, Thailand, March 19–21, 2014, 1–4. doi:10.1109/iEECON.2014.6925970

Cotilla-Sanchez, E., Hines, P. D. H., Barrows, C., Blumsack, S., and Patel, M. (2013). Multi-attribute Partitioning of Power Networks Based on Electrical Distance. *IEEE Trans. Power Syst.* 28, 4979–4987. doi:10.1109/TPWRS.2013.2263886

Holttinen, H., Groom, A., Kennedy, E., Woodfin, D., Barroso, L., Orths, A., et al. (2021). Variable Renewable Energy Integration: Status Around the World. *IEEE Power Energy Mag.* 19, 86–96. doi:10.1109/MPE.2021.3104156

Hung, D. Q., Mithulananthan, N., and Lee, K. Y. (2014). Determining PV Penetration for Distribution Systems with Time-Varying Load Models. *IEEE Trans. Power Syst.* 29, 3048–3057. doi:10.1109/TPWRS.2014.2314133

Lee, J. R., Gharan, S. O., and Trevisan, L. (2014). Multiway Spectral Partitioning and Higher-Order Cheeger Inequalities. *J. ACM* 61, 1–30. doi:10.1145/2665063

Leeuwen, T. v., and Moser, A. (2017). "Impact of Flexible Transmission Assets on Day-To-Day Transmission Grid Operation under Uncertainties," in International ETG Congress 2017, Bonn, Germany, November 28–29, 2017, 1–6.

Li, Z., Cheng, Z., Si, J., Zhang, S., Dong, L., Li, S., et al. (2021). Adaptive Power Point Tracking Control of PV System for Primary Frequency Regulation of AC Microgrid with High PV Integration. *IEEE Trans. Power Syst.* 36, 3129–3141. doi:10.1109/TPWRS.2021.3049616

Lin, X., Shu, T., Tang, J., Ponci, F., Monti, A., and Li, W. (2022). Application of Joint Raw Moments-Based Probabilistic Power Flow Analysis for Hybrid AC/VSC-MTDC Power Systems. *IEEE Trans. Power Syst.* 37, 1399–1412. doi:10.1109/TPWRS.2021.3104664

Ma, D., Hu, X., Zhang, H., Sun, Q., and Xie, X. (2021). A Hierarchical Event Detection Method Based on Spectral Theory of Multidimensional Matrix for Power System. *IEEE Trans. Syst. Man. Cybern. Syst.* 51, 2173–2186. doi:10.1109/TSMC.2019.2931316

Park, Y., and Kim, J. (2006). A Survey on the Power System Modeling Using a Clustering Algorithm. *Proc. KIEE Conf.*, 410–411.

Sanchez-Garcia, R. J., Fennelly, M., Norris, S., Wright, N., Niblo, G., Brodzki, J., et al. (2014). Hierarchical Spectral Clustering of Power Grids. *IEEE Trans. Power Syst.* 29, 2229–2237. doi:10.1109/TPWRS.2014.2306756

Sarajpoor, N., Rakai, L., Arteaga, J., and Zareipour, H. (2021). A Shape-Based Clustering Framework for Time Aggregation in the Presence of Variable Generation and Energy Storage. *IEEE Open J. Power Energy* 8, 448–459. doi:10.1109/OAJPE.2021.3097366

Sehsalar, O. Z., Galvani, S., and Farsadi, M. (2019). New Approach for the Probabilistic Power Flow of Distribution Systems Based on Data Clustering. *IET Renew. Power Gener.* 13, 2531–2540. doi:10.1049/iet-rpg.2018.6264

Shaheen, M. A. M., Hasanien, H. M., Mekhamer, S. F., and Talaat, H. E. A. (2019). Optimal Power Flow of Power Systems Including Distributed Generation Units Using Sunflower Optimization Algorithm. *IEEE Access* 7, 109289–109300. doi:10.1109/ACCESS.2019.2933489

Si, C., Xu, S., Wan, C., Chen, D., Cui, W., and Zhao, J. (2021). Electric Load Clustering in Smart Grid: Methodologies, Applications, and Future Trends. *J. Mod. Power Syst. Clean. Energy* 9, 237–252. doi:10.35833/MPCE.2020.000472

Teng, J.-H., Luan, S.-W., Lee, D.-J., and Huang, Y.-Q. (2013). Optimal Charging/discharging Scheduling of Battery Storage Systems for Distribution Systems Interconnected with Sizeable PV Generation Systems. *IEEE Trans. Power Syst.* 28, 1425–1433. doi:10.1109/TPWRS.2012.2230276

The International Renewable Energy Agency (2021). Available at: https://www.irena.org/Statistics/View-Data-by-Topic/Capacity-and-Generation/Statistics-Time-Series (Accessed March 22, 2022).

Tyuryukanov, I., Popov, M., van der Meijden, M. A. M. M., and Terzija, V. (2018). Discovering Clusters in Power Networks from Orthogonal Structure of Spectral Embedding. *IEEE Trans. Power Syst.* 33, 6441–6451. doi:10.1109/TPWRS.2018.2854962

von Luxburg, U. (2007). A Tutorial on Spectral Clustering. *Stat. Comput.* 17, 395–416. doi:10.1007/s11222-007-9033-z

von Luxburg, U., Belkin, M., and Bousquet, O. (2008). Consistency of Spectral Clustering. *Ann. Stat.* 36. doi:10.1214/009053607000000640

Wang, S., Su, H., and Zhang, Z. (2021). "Probabilistic Power Flow Calculation Considering the Correlation of Distributed Power Sources," in 2021 IEEE

International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), Dialan, China, August 27–28, 2021, 53–58. doi:10.1109/AEECA52519.2021.9574164

Wiget, R., Vrakopoulou, M., and Andersson, G. (2014). "Probabilistic Security Constrained Optimal Power Flow for a Mixed HVAC and HVDC Grid with Stochastic Infeed," in 2014 Power Systems Computation Conference, Wroclaw, Poland, August 18–22, 2014, 1–7. doi:10.1109/PSCC.2014.7038408

Zhu, X., Liu, C., Su, C., and Liu, J. (2020). Learning-based Probabilistic Power Flow Calculation Considering the Correlation Among Multiple Wind Farms. *IEEE Access* 8, 136782–136793. doi:10.1109/ACCESS.2020.3011511

Zimmerman, R. D., and Murillo-Sanchez, C. E. (2020). *MATPOWER User's Manual Version 7.1*. Madison, WI: Power Systems Engineering Research Center.

Conflict of Interest: JL was employed by the company Korea Electric Power Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

NOMENCLATURE

Abbreviations

PV Photovoltaic

VRE Variable renewable energy

Symbols

$\partial(C)$ Boundary of cluster C

γ Eigengap

λ Eigenvalue

$\phi(C)$ Expansion of cluster C

ϕ_{max} Maximum expansion of clusters

e Mean expansion of clusters

μ Mean of random variable

α, β Parameter of f_{beta}

ρ_{max} Probabilistic clustering evaluation index based on expansion max

ρ_{mean} Probabilistic clustering evaluation index based on expansion mean

σ Standard deviation of random variable

δ_l Voltage angle at bus k

$vol(C)$ Volume of cluster C

$\mathbf{B_{kl}}$ Susceptance of the admittance Y_{kl}

E Edge set

$\boldsymbol{e}$ Normalized coordinates of bus i

$\boldsymbol{f_{beta}}$ Beta probability density function

FC Total fuel cost

FF Fill factor

G Graph

$\mathbf{G_{kl}}$ Conductance of the admittance Y_{kl}

$\mathbf{I_{SC}}$ Short-circuit current

$\mathbf{K_c}$ Voltage temperature coefficients

$\mathbf{K_i}$ Current temperature coefficients

L Laplacian matrix

$\mathbf{L_n}$ Normalized Laplacian matrix

$\mathbf{n_g}$ Total number of generator buses

NI Neutrality index

$\mathbf{n_l}$ Total number of load buses

$\mathbf{n_{PV}}$ Total number of PV buses

$\mathbf{P_{Ge}}$ Active power generation at generation bus i

$\mathbf{P_{inj,k}}$ Active power injected into bus k

$\mathbf{P_{Lj}}$ Active power consumption at load bus j

$\mathbf{P_{loss}}$ Power loss in the network

$\mathbf{P_{PVk}}$ PV generation at PV bus k

$\mathbf{Q_{inj,k}}$ Reactive power injected into bus k

s Random variable of solar irradiance

$\mathbf{T_a}$ Ambient temperature of PV cell

$\mathbf{T_c}$ Cell temperature of PV cell

$\mathbf{T_n}$ Nominal operating temperature of PV cell

$\boldsymbol{v}$ Eigenvector

V Vertex set

$\mathbf{V_{oc}}$ Open-circuit voltage

$\boldsymbol{x_i}$ Coordinates of bus i

$\mathbf{Y_{kl}}$ Admittance of branch from buses from i to j