Sharma *et al. Comput Soc Netw* (2019) 6:4
https://doi.org/10.1186/s40649-019-0064-3

Computational Social Networks

RESEARCH Open Access

A network science-based *k*-means++ clustering method for power systems network equivalence

Dhruv Sharma[1*], Krishnaiya Thulasiraman[2], Di Wu[3] and John N. Jiang[1]

*Correspondence:
dhruv.sharma@ou.edu
[1] Department of Electrical and Computer Engineering, University of Oklahoma, Norman, USA
Full list of author information is available at the end of the article

Abstract

Network equivalence is a technique useful for many areas including power systems. In many power system analyses, generation shift factor (GSF)-based bus clustering methods have been widely used to reduce the complexity of the equivalencing problem. GSF captures power flow on a line when power is injected at a node using bus to bus electrical distance. A more appropriate measure is the one which captures what may be called the electrical line distance with respect to a bus termed as relative *bus to line* distance. With increase in power transactions across different regions, the use of relative *bus to line* distance becomes appropriate for many analyses. Inspired by the recent trends in network science on the study of network dynamics based on the topological characteristics of a network, in this paper, we present a bus clustering method based on average electrical distance (AED). AED is independent of changes in location of slack bus and is based on the concept of electrical distance introduced in the context of molecular chemistry and pursued later for applications in social and complex networks. AED represents the AED from a bus to buses of the transmission line of interest. We first propose an AED-based method to group the buses into clusters for power systems network equivalence using *k*-means clustering algorithm integrated with silhouette analysis. One limitation of this method is that despite its speed, sometimes it may yield clusters of inferior quality compared to the optimal solution. To overcome this limitation, we next present our improved clustering method which incorporates a seeding technique that initializes centroids probabilistically. We also incorporate a technique in our method to find the number of clusters, *k*, to be given as input to our clustering algorithm. The resulting algorithm called AED-based *k*-means++ clustering method yields a clustering that is *O(logk)* competitive. Our network equivalence technique is next described. Finally, the efficacy of our new equivalencing technique is demonstrated by evaluating its performance on the IEEE 300-bus system and comparing that to the performance of our AED-based method (Sharma et al. in Power network equivalents: a network science-based *k*-means clustering method integrated with silhouette analysis. In: Complex networks and their application VI, Lyon, France. p. 78–89, 2017) and the existing GSF-based method.

Keywords: Average electrical distance, Power systems network equivalence, Generation shift factor, *k*-means++ algorithm

Introduction

The past few decades have witnessed a new movement of interest and research in the study of complex networks, i.e., networks whose structure is irregular, complex and dynamically evolving in time such as power grids, communication networks, biological networks, social networks, etc. Investigating dynamics in such complex networks requires an understanding of the interaction between network topology and specific domain constraints. For example, the study of power grids requires basic circuit laws, relating voltages and currents, to be incorporated along with the network topology. In this paper, we explore this interaction in the context of deriving simplified equivalent networks as representations of large power grids. In this paper, the terms power systems, power system network, and power network are used interchangeably. Power network equivalencing has been studied in the literature with different objectives. Our objective is to study the equivalents appropriate for electrical market analysis.

Electrical market analysis involving power exchange is becoming more and more complex due to the size and degree of interconnections in modern power systems due to economic, political, and environmental reasons [1]. With such inherent complexities and information deficiency, it is difficult for participants to make operational and market decisions at buses that are sensitive to the condition of transmission lines. For these reasons, it is necessary to develop an enhanced method to support decisions, particularly, those sensitive to major and critical transmission lines. The analysis can be computationally challenging, especially, when a full AC implementation approach is used [2]. Compared to the full AC analysis, a simplified analysis of the network can be done by means of full network DC power flow model [3]. Although the full AC analysis would be the most accurate approach, the DC approach allows network operators to make informed dispatch decisions thereby saving time and effort required. The enormous size of power system networks makes full network DC analysis computationally taxing. To reduce the computational burden and to simplify the analysis of electricity markets, several network equivalence models have been used [2, 4–6].

Various approaches for network equivalencing have been presented in [7–12]. These approaches follow the traditional method of eliminating less important elements from the system on the basis of geographic and electric parameters. Such elimination results in partitioning of the network into three clusters of buses: a cluster of internal buses, a cluster of external buses, and a cluster of boundary buses that divide the external buses from the internal buses. Due to their trivial impact on the internal system, remote generators and transmission lines connected to the boundary buses may be eliminated with minor impact on decisions. However, irrespective of the bus demarcation, market analysis of large power system networks requires retaining the desired buyer/seller pairs corresponding to operating zones in order to understand the impact of power flows on transmission lines from the buyer/seller pairs at various buses. Hence, in this study, we do not eliminate the external buses but rather focus on the line flows between various operating areas. These line flows are called tie-line flows.

In this paper, the metric developed for the study has properties similar to resistance distance which has been widely used since the early days of electrical circuit theory. There are several books on the basics of circuit analysis that deal with resistance distance and topological formula for resistance distance, for example [13]. However, recently this

concept gained increased importance in view of its applications in areas outside electrical circuit theory [14–20].

In recent studies, network equivalencing has been done using generation shift factor (GSF)-based methods [5, 6, 21, 22]. In these methods, buses with similar impact on the interconnecting tie-line flows evaluated using GSFs are grouped together. In order to improve the efficiency and accuracy of bus clustering, [6] use is made of *k*-means algorithm based on GSF to cluster the buses. However, GSFs are sensitive to the change in the location of slack generator. Network operators supervising different regions of the interconnection might not be aware of the slack bus change, and hence there could be discrepancies in decision making, which can provoke an impact on regional power transactions. Therefore, our main objective in this paper is to develop a new clustering method as well as new network equivalent that overcomes the limitation of the GSF-based methods.

Our work in this paper includes our previous work [23] and its enhancements. The rest of this paper is organized as follows. "Preliminaries" section gives an introduction to all the basic concepts that are of interest in the development of techniques discussed in subsequent sections. We introduce, in "Average electrical distance" section, the concept of average electrical distance (AED) and discuss a relationship between AED and GSF. Also, the relevance of our work in the context of social network analysis is given in "Implications and relevance for social network analysis" section. "*k*-means algorithm" section gives an introduction to the *k*-means algorithm for clustering, followed by "AED-based *k*-means bus clustering method" section presenting our proposed AED-based *k*-means clustering method, which, however, has certain limitations. In "AED-based *k*-means++ bus clustering method" section, we present our improved AED-based *k*-means++ clustering method, to overcome the said limitations. The new method uses a seeding technique [24] for initialization of centroids. We call this augmented algorithm as AED-based *k*-means++ algorithm. We also incorporate in this method silhouette analysis to determine the number *k* of clusters to be given as input to the clustering method. The resulting AED-based *k*-means++ method yields clustering which is *O(logk)* competitive. "Power network equivalencing based on AED-based *k*-means++ clustering method" section presents how we use clustering to develop a power network equivalent suitable for market analysis. To demonstrate the efficacy of our approach for clustering and equivalencing, the rest of "Power network equivalencing based on AED-based *k*-means++ clustering method" section is concerned with experimental and comparative evaluations of our network equivalence method applied on the 300-bus system. We conclude this paper with "Conclusion" section, highlighting the main findings of this work, their implications, and our continuing attempt to find an efficient network equivalencing technique for power systems and its implications in the context of social network analysis.

Preliminaries

In this section, we introduce the basic concepts of Laplacian matrix of a graph, electrical distance, and generation shift factor.

Laplacian matrix of a graph

Consider a graph $G = (V, E)$ with vertex set $V = \{0, 1, \ldots, n\}$. Edge $e \in E$ connecting vertices i and j is denoted by (i, j). We assume there are no loops on any vertices and there are no parallel edges connecting the vertices. In this paper, the terms, vertices and nodes, as well as links and edges, will be used interchangeably. Let an edge (i, j) be assigned a weight w_{ij}, a positive real number. If there is no edge connecting i and j then $w_{ij} = 0$. Two vertices i and j are adjacent if there is an edge (i, j). A vertex j is incident on vertex i if there is an edge connecting i and j. The degree of a vertex i denoted by $\deg(i)$ is the sum of the weights of the edges incident on i.

The Laplacian matrix, $Y = [y_{ij}]$ of G is an $(n + 1) \times (n + 1)$ matrix defined as follows:

$$y_{ij} = \begin{cases} -w_{ij}, & \text{if } i \neq j \\ \deg(i), & \text{if } i = j \end{cases} \tag{1}$$

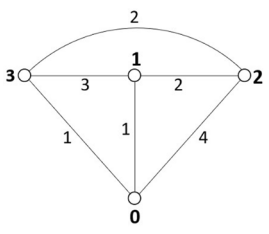
As an example, a graph G and its adjacent matrix are shown in Fig. 1a, b. Note that $y_{ij} = y_{ji}$. It can be seen that the sum of all the elements in any row and the sum of all the elements in any column are both equal to zero. So, Y is singular and has no inverse. To handle this singularity problem of Y, two different approaches are adopted, namely, the eigenvalue approach and the determinant approach.

In the eigenvalue approach, the pseudo-inverse Y^+ of Y is used. The properties of G are studied in terms of the elements of Y^+. This approach is quite popular among mathematicians [25, 26]. In this paper, we follow the determinant approach which is popular in the electrical engineering community. In this approach, we first remove a row and the corresponding column from the Laplacian matrix, Y. Let us assume that the vertex labelled o called the datum node or slack node is removed. The resulting matrix denoted by $Y(o)$ is called a reduced Laplacian matrix. The reduced Laplacian matrix $Y(o)$ of Y in Fig. 1b is shown in Fig. 1c. It can be shown that the matrix $Y(o)$ is non-singular and it has several other properties, for example, [13].

In all discussions in this paper, we will use $Y = [y_{ij}]$ to denote the reduced Laplacian of the network. We wish to draw attention to Cetinay et al. [27] where the authors investigate the impact of topology on power flow using spectral graph theory.

Electrical distance

Consider an electrical network N represented by the graph G. The admittance of the edge (i, j) in N serves as the weight w_{ij} in G. Each edge (s, t) in G is associated with two variables, voltage, v_{st} and current i_{st}, Fig. 2. Then by Ohm's law, we get



a

$$Y = \begin{array}{c} \\ \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \\ \mathbf{0} \end{array} \begin{array}{c} \begin{array}{cccc} \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{0} \end{array} \\ \begin{bmatrix} 6 & -2 & -3 & -1 \\ -2 & 8 & -2 & -4 \\ -3 & -2 & 6 & -1 \\ -1 & -4 & -1 & 6 \end{bmatrix} \end{array}$$

b

$$Y(o) = \begin{array}{c} \\ \mathbf{1} \\ \mathbf{2} \\ \mathbf{3} \end{array} \begin{array}{c} \begin{array}{ccc} \mathbf{1} & \mathbf{2} & \mathbf{3} \end{array} \\ \begin{bmatrix} 6 & -2 & -3 \\ -2 & 8 & -2 \\ -3 & -2 & 6 \end{bmatrix} \end{array}$$

c

Fig. 1 **a** Graph G with weighted edges; **b** Laplacian Matrix of G; **c** reduced Laplacian Matrix, $Y(o)$

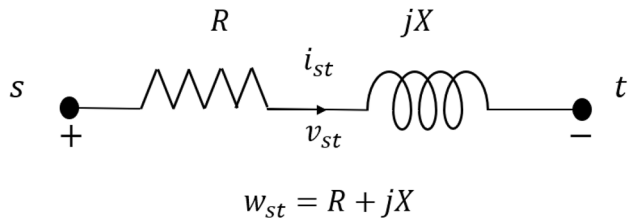


Fig. 2 Demonstration of Ohm's law

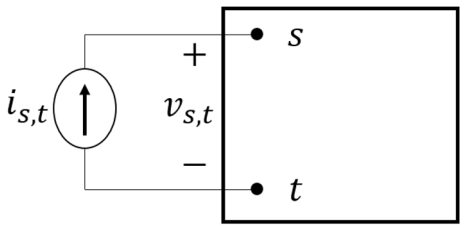


Fig. 3 Current source across the nodes (s, t)

$$\frac{i_{st}}{v_{st}} = w_{st}, \quad \text{the admittance of } (s,t). \tag{2}$$

In the electrical engineering literature, the reduced Laplacian matrix, Y is called the node-to-datum matrix of N with vertex o as the datum vertex. Datum vertex is also known as the slack vertex. Let $Z = [Z_{ij}]$ be the inverse of the reduced Laplacian matrix. This matrix is called Z-bus (Z_{bus}) matrix in the power engineering literature. Also, the ground is usually used as the slack vertex.

Suppose we now connect a current source of value $i_{s,t}$ across nodes (s, t), Fig. 3. Let $v_{s,t}$ be the voltage across s and t. Then the electrical distance between s and t, denoted as $r_{s,t}$ is defined as

$$r_{s,t} = \frac{v_{s,t}}{i_{s,t}}. \tag{3}$$

We now proceed to show how to evaluate $r_{s,t}$ across all pairs of nodes s and t. Note that definition of $r_{s,t}$ does not require that s and t be connected by edge (s, t) in the network N. Suppose we now connect a current source of value i_s across the vertices o and for each $s \in \{1, 2, \ldots, n\}$, as shown in Fig. 4. Note that the current flows from node o to node s. Let v_s denote the voltage from node s to node o. If we let $V = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ and $I = \begin{bmatrix} i_1 \\ \vdots \\ i_n \end{bmatrix}$, then the matrix Z represents the relation between V and I as

$$\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = Z \begin{bmatrix} i_1 \\ \vdots \\ i_n \end{bmatrix} \tag{4}$$

Therefore,

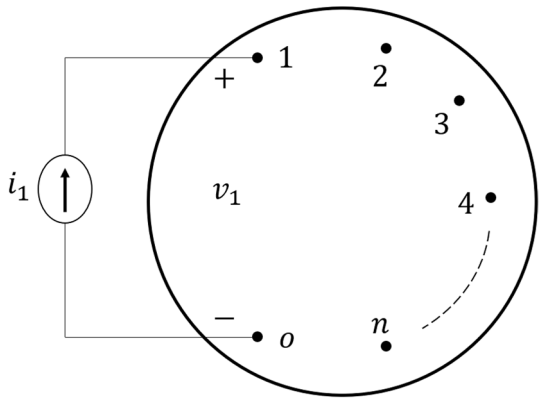


Fig. 4 Current source connected across $(0, 1), (0, 2), \ldots, (0, n)$

$$Z_{ss} = \left.\frac{v_s}{i_s}\right|_{i_j=0, j\neq s} \tag{5}$$

$$Z_{st} = \left.\frac{v_s}{i_s}\right|_{i_j=0, j\neq t} \tag{6}$$

Note that Z_{ij} is a complex number denoted by $(R_{ij} + jX_{ij})$.

It follows from (5) that $r_{s,o} = Z_{ss}$, for all s. It can be shown that for all i and j

$$r_{ij} = Z_{ii} + Z_{jj} - 2Z_{ij}. \tag{7}$$

In the power engineering literature, electrical distance r_{st} is called Thevenin resistance/impedance between the nodes s and t, denoted by $Z_{th,st}$. Thus,

$$Z_{th,st} = r_{st} = Z_{ss} + Z_{tt} - 2Z_{st}. \tag{8}$$

Also, in the chemistry literature, the electrical distance is referred to as resistance distance [14].

Generation shift factor

Given a power network with two or more interconnected areas, in power market analysis, a simpler equivalent network is needed which preserves the flows across the lines (edges) connecting different areas. In this context, the concept of GSF [5] was introduced and used in determining power network equivalents.

Let us consider a network N represented by graph G. Consider a line connecting two nodes u and v. Suppose we inject a current of unit value at node i i.e., connect a current source of unit value between node i and the slack node o. Then the current that flows through the line (u, v) is called the generation shift factor of i with respect to (u, v), denoted as $g_{uv,i}$. To find a formula for $g_{uv,i}$ consider (4). Then,

$$v_u = z_{u1}i_1 + z_{u2}i_2 + \cdots + z_{ui}i_i + \cdots + z_{uu}i_u, \tag{9}$$

$$v_v = z_{v1}i_1 + z_{v2}i_2 + \cdots + z_{vi}i_i + \cdots + z_{vv}i_v. \tag{10}$$

If $i_j = 0$ for all $j \neq i$ and $i_i = 1$, then we get $v_u = z_{ui}$ and $v_v = z_{vi}$. Then the current flowing through line (u, v) is given by

$$i_{uv} = \frac{z_{ui} - z_{vi}}{c_{uv}}, \tag{11}$$

where c_{uv} is the admittance of line (u, v). Thus, the GSF of i with respect to (u, v) is

$$g_{uv,i} = \frac{z_{ui} - z_{vi}}{c_{uv}}. \tag{12}$$

Average electrical distance

Average electrical distance (AED) is defined as the electrical distance of a specific bus with respect to a transmission line or a tie-line connecting different operating zones. It can also be viewed as the relative electrical distance of a bus w.r.t. a line. The AED, $d_{uv,i}$, is defined as

$$d_{uv,i} = \left| \frac{Z_{th,ui} - Z_{th,vi}}{2} \right|, \tag{13}$$

where $Z_{th,ui}$ is the electrical distance between buses u and i, Fig 5. Using (8), we get

$$d_{uv,i} = \left| \frac{(Z_{uu} - 2Z_{ui} + Z_{ii}) - (Z_{vv} - 2Z_{vi} + Z_{ii})}{2} \right|, \tag{14}$$

$$= \left| \frac{\bar{Z}_{uu} - \bar{Z}_{vv}}{2} - (\bar{Z}_{ui} - \bar{Z}_{vi}) \right|, \tag{15}$$

where Z_{ui} is the corresponding (u, i) element in the bus impedance matrix shown in (16).

$$Z = \begin{bmatrix} Z_{11} & \cdots & Z_{1i} & \cdots & Z_{1n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{u1} & \cdots & Z_{ui} & \cdots & Z_{un} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ Z_{n1} & \cdots & Z_{ni} & \cdots & Z_{nn} \end{bmatrix}. \tag{16}$$

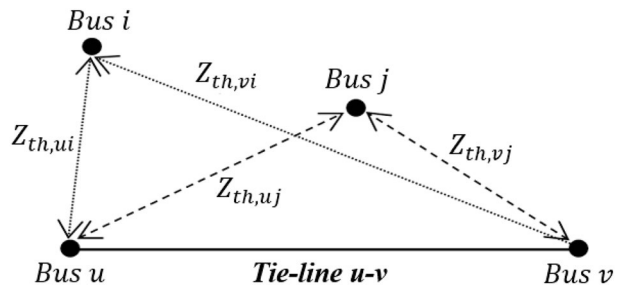


Fig. 5 Illustration of average electrical distance

In (16), the diagonal elements of the impedance matrix represent the Thevenin impedance at each corresponding bus. For real-time analysis, it is required to execute calculations of electrical distance at short intervals which can be achieved through simplification of electrical distances. Since transmission lines usually have a high X/R ratio (Fig. 2), the resistances of the lines can be neglected compared to the high reactances of lines. Thus, the impedances in (14) can be simplified as shown in (17).

$$X_{th,ui} = \bar{X}_{uu} - 2\bar{X}_{ui} + \bar{X}_{ii}, \tag{17}$$

where elements in $\bar{X} = [\bar{X}_{ij}]$ represent the approximation of elements in Z, i.e., $Z_{ij} \approx \bar{X}_{ij}$. Then, AED defined in (13) can be rewritten as

$$d_{uv,i} = \left| \frac{X_{th,ui} - X_{th,vi}}{2} \right| = \left| \frac{\bar{X}_{uu} - \bar{X}_{vv}}{2} - (\bar{X}_{ui} - \bar{X}_{vi}) \right|. \tag{18}$$

In previous studies [6, 21], GSF-based bus clustering methods have been used to analyse the impact of injections from groups of buses on the power flows of different transmission lines. In these GSF-based methods, buses that have similar contributions to the power flows on lines of interest are grouped together based on GSFs. However, GSF does not take into account the impact of injection on buses u and v. As shown below, AED is indeed a measure that captures the impact of injection at u and v. This important characteristic of AED can be explained by analysing the relationship between AED and GSF. GSFs can be calculated using elements of $\bar{X}$ as shown in (19).

$$g_{uv,i} = \frac{\bar{X}_{ui} - \bar{X}_{vi}}{x_{uv}}. \tag{19}$$

Similarly, the sum of GSFs of bus u and bus v with respect to tie-line uv is given by

$$g_{uv,u} + g_{uv,v} = \left| \frac{\bar{X}_{uu} - \bar{X}_{vv}}{x_{uv}} \right|. \tag{20}$$

Comparing (18–20), we can see that $d_{uv,i}$ enhances $g_{uv,i}$ by adding an additional term capturing the impact of injection at buses, u and v, on the power flow across the tie-line, uv.

In the rest of the paper, we discuss our improved clustering method based on AED and compare the results with the existing GSF-based clustering method.

Calculation of AED

Algorithm given below to determine AED for each bus in the network with respect to the tie-lines of interest includes the following main steps:

Algorithm 1: Calculation of AED

1. *Creation of bus impedance matrix* According to the system data, bus admittance matrix is first developed. Then, bus impedance matrix, shown in (16), is created by calculating the inverse of bus admittance matrix.
2. *Calculation of Thevenin impedance* Using the elements of bus impedance matrix obtained in Step 1, the Thevenin impedances for each pair of buses are calculated according to (17).

3. *Calculation of AED* After calculating the Thevenin impedances, AEDs between buses and tie-lines of interest are calculated using (18). The results of AEDs are used to create a matrix as shown in Fig. 6. In this matrix, each row corresponds to a tie-line and each column corresponds to a bus in the system. Thus, an element in the matrix corresponds to AED from a bus to a tie-line in the system.

□

Implications and relevance for social network analysis

In the graphical representation of a social network, link weights are all unity. Different types of metrics/measures are defined to determine certain properties of links or nodes. For example, a measure called PageRank is used to rank the nodes in terms of their importance [28]. As another example, the betweenness measure of a link is used to determine the importance of a link. A link e is considered more important than another link e', if the fraction of the total number of messages that flow through e is greater than that for e'. In view of the connection between random walk and current flow in a resistance network [18], this fraction is in fact equal to the current through the link when a unit current is injected at a node. Taking advantage of this connection in [15], a measure similar to GSF is used to determine the betweenness measures of links. A detailed discussion of many of these measures is given in [29].

The concept of role discovery in networks was first studied in sociology [30, 31]. In this context, roles considered are social roles. Thus, role discovery has become an important topic in social network analysis. Recently, role discovery has been studied in other settings such as online social networks, technological networks, biological networks, web graph, etc.

In [32], a comprehensive review of literature on role discovery in network has been given. This paper discusses the problem of identifying clusters in a network such that all nodes in each cluster are equivalent in some sense. Two types of equivalence are considered: graph-based equivalence and feature-based equivalence. Several challenges that arise in the application of role discovery in non-static network such as dynamic and streaming graphs are also discussed in [32].

Our work in this paper is about clustering in a power network and its application in deriving a simplified approximate equivalent network that preserves flows along certain lines. The GSF and AED are measures that are defined for each node with respect

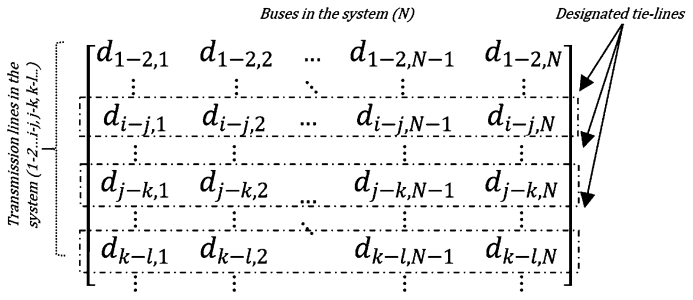


Fig. 6 Illustration of AED matrix

to certain lines. If we set the line weights to unity, then these measures in the context of social networks capture the fraction of total messages that flow through a link when messages arrive (or injected) at a node. Therefore, the work presented in this paper is relevant to social network studies. For example, a problem of interest is to determine clusters such that the total flow carried by inter-cluster links is optimized. Once such clusters are identified, we can determine simplified approximate equivalent network as explained in "Power network equivalencing based on AED-based *k*-means++ clustering method" section that can be used to predict the flows across the clusters. Further discussion of these ideas is given in "Conclusion" section.

k-means algorithm

k-means algorithm is one of the most popular clustering techniques in unsupervised learning tasks. Given a set of nodes or buses, this algorithm has been efficiently used to partition a network into k clusters [33]. This is based on the optimal placement of centroid for the respective cluster in a network [34].

In this algorithm, initially the network is divided into k clusters with each cluster defined by a reference bus (centroid). Remaining buses are then partitioned and assigned appropriately to the clusters based on the closeness of each bus to k reference buses. Then cluster adjustments are made with the calculation of new centroids. These centroids act as new reference points for the next partitioning of all the buses. These adjustments naturally produce error minimum of which corresponds to "Voronoi configuration" [35] which results in reference locations at the centroid of the clusters. The error measure or potential function is the sum of all the variances and is given as shown in (21).

$$\phi = \min \sum_{j=1}^{k} \sum_{i=1}^{n_j} |x_{ij} - \mu_j|^2, \tag{21}$$

where k is the total number of clusters; n_j is the number of buses belonging to the jth cluster; x_{ij} represents ith bus in the jth cluster; μ_j is the centroid in the jth cluster and the term; and $|x_{ij} - \mu_j|$ represents the distance between x_{ij} and μ_j.

The process becomes iterative in order for the clusters to reach a local minimum which is dependent on the initial selection of the reference buses. The *k*-means algorithm keeps on adjusting the centroids after each partition making it more dynamic to the changes. The *k*-means algorithm is explained in Algorithm 2.

Algorithm 2: k-means algorithm

1. *Selecting initial cluster centroids* Cluster formation is initialized by selecting k centroids, i.e., $\mu_1, \mu_2, \ldots, \mu_k$ in the network. These centroids act as initial reference points for the buses to be assigned to an appropriate cluster.
2. *Grouping buses into clusters* For the selected k centroids, based on the Euclidean distance, a bus is assigned to a cluster which has its centroid closest to the bus.
3. *Recalculating centroid positions* After all the buses are assigned to respective clusters, the new centroid of each cluster is recalculated as shown in (22).

$$\mu_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij} \tag{22}$$

4. *Evaluating objective function in* (21) After all buses are grouped into the clusters in Step 2, the potential function in (21) is evaluated.
5. *Iterations of algorithm* Steps 2–4 are repeated until the centroid of each cluster ceases to change its position with further iterations.

□

AED-based *k*-means bus clustering method

In this section, we discuss our AED-based improved clustering method that uses *k*-means algorithm for power system network equivalence. This method uses a simple iterative technique known as Lloyd's algorithm for finding a locally minimal solution [34]. Further, it utilizes AED as a measure of distance between the buses in the system. Integration of AED makes *k*-means algorithm more relevant to the power system network study. This is because AED gives a measure of distance of a bus with respect to a tie-line. This algorithm proves to be sufficiently accurate for the independent analysis done by various utilities on their networks.

AED-based *k*-means algorithm

With reference to the power system network equivalence, it is preferred to incorporate a measure that can reflect the true distance from a power system network perspective. AED presented in "Average electrical distance" section is one such measure. In the network, while determining the clusters, we replace the actual distance term in the variance calculation with the AED measure. Specifically, in (21), we replace the location of the bus x_{ij} with $d(x_{ij})$ which represents the AED of the ith bus in the jth cluster with respect to set of tie-lines and $d(\mu_j)$ represents the AED-based measure of the centroid in the jth cluster with respect to the tie-lines. If the number of tie-lines is greater than 1, $d(x_{ij})$ is replaced with the average of AEDs considered for the given set of tie-lines. The updated potential function is shown in (23).

$$\phi = \min \sum_{j=1}^{k} \sum_{i=1}^{n_j} |d(x_{ij}) - d(\mu_j)|^2. \tag{23}$$

This improves the *k*-means algorithm when applied in power system setting compared to general distance used in the classical algorithm.

AED-based *k*-means++ bus clustering method

In this section, we introduce AED-based improved *k*-means++ algorithm which can be used for clustering of large power system networks.

k-means++ algorithm

Although the results showed improvements compared to other methods used for clustering power system networks, the AED-based improved *k*-means clustering method may deliver inconsistent results for large power system networks comprising a large dataset. This is due to the fact that *k*-means algorithm uses random centroids for initialization and thus achieves different results for each simulation. To address this problem, a seeding technique was proposed in [24] that selected the first centroid position at random and then initialized the remaining centroids by sampling probabilistically, proportional to the squared distance of the nearest centroid. This made an improvement in the *k*-means algorithm which helps to achieve a clustering which is $O(\log k)$ competitive. This is achieved by considering a potential function that satisfies the following relation for any set of buses:

$$\mathbb{E}[\phi] \leq 8(\ln k + 2)\phi^*, \tag{24}$$

where ϕ^* is the potential function corresponding to the set of cluster centroids in the network. The resulting augmented *k*-means algorithm is called *k*-means++ algorithm. In our study, we modify our previous AED-based *k*-means algorithm by incorporating the seeding technique to obtain a much improved AED-based *k*-means++ algorithm for bus clustering.

We would like to point out that *k*-means++ algorithm, due to its initialization process, produces starting centroids uniformly distributed for different iterations compared to *k*-means algorithm starting centroids. This is illustrated in [36].

AED-based *k*-means++ algorithm

Different from commonly used *k*-means++ algorithm, the improved AED-based algorithm groups the buses in a power system into various clusters based on the closeness between buses and the centroid of each cluster in terms of AEDs. The objective of the improved AED-based *k*-means++ algorithm is also to minimize the same potential function, ϕ, as shown in (23).

To achieve the objective described in (23) with the initial seeding as shown in (24), the improved AED-based algorithm, shown in Algorithm 3, includes the following main steps:

Algorithm 3: AED-based k-means++ algorithm

1. *Cluster initialization* Cluster formation is initialized by selecting one centroid μ_1, chosen uniformly at random in the network. This centroid acts as initial reference point for the buses to be assigned to an appropriate cluster. Those buses that are closer to μ_1 are later assigned to the same cluster.
2. *Determining cluster centroids* Given the centroids, $\mu_1, \ldots, \mu_{j-1}$, a new centroid, μ_j, is chosen and each bus, i, is selected with probability

$$\frac{D(i)^2}{\sum_{i=1}^{n} D(i)^2},$$

where $D(i) = \min|d(uv, i) - d(uv, \mu_r)|$, where $r = 1, \ldots, j-1$.

3. Step 2 is repeated until we get all the centroids.
4. *k-means algorithm* Proceed as with the standard k-means algorithm (Algorithm 2) with AED used as distance measure as discussed in "AED-based k-means bus clustering method" section.

□

Silhouette value analysis

An important problem in the application of k-means algorithm is to determine appropriate value of k. This problem has been extensively studied in the mathematical statistics literature [37, 38]. The authors in [39] identified certain best performing methods. In [37], authors determined through extensive simulation studies that all the best performers do quite well in selecting the appropriate number of clusters to be selected. In our work, we use the silhouette value analysis proposed in [40] that is also among the best performers.

Silhouette value analysis is a graphical partitioning technique [41] allowing an appreciation of the relative quality of clusters. In our study, the silhouette value analysis is used to enhance the quality of clusters identified by the improved AED-based k-means++ clustering algorithm. The main steps of the silhouette value analysis, which are incorporated into the improved AED-based k-means algorithm are explained below:

Algorithm 4: Silhouette value analysis algorithm

Consider that in a network with k clusters, a bus $i \in \{1, 2, \ldots, n\}$ is assigned to a cluster $j \in \{1, 2, \ldots, k\}$. Let n_j be the number of buses in cluster j and k is the arbitrarily selected number of clusters. Clusters neighbouring j are represented by $m \in \{1, 2, \ldots, k-1\}$ such that $m \neq j$; these are all the clusters other than j.

1. *Evaluating closeness between buses in a cluster* In this step, the average closeness between the buses in cluster j is evaluated by calculating AED measure, ac_{ij}, with respect to the tie-lines in the network. This is shown in (25).

$$ac_{ij} = \frac{1}{n} \sum_{\substack{r=1 \\ r \neq i}}^{n_j} |d(uv, i) - d(uv, r)| \tag{25}$$

2. *Evaluating closeness between each bus and clusters* In this step, the minimum of average closeness of a bus i with respect to each cluster $m \neq j$ is evaluated which is given by eb_{im}.

$$eb_{im} = \frac{1}{n_m} \sum_{r=1}^{n_m} |d(uv, i) - d(uv, r)|, \quad m = 1, \ldots, k-1; \quad m \neq j \tag{26}$$

where n_m is the number of buses in cluster m.

3. *Calculating average silhouette coefficient of all the clusters* The silhouette coefficient of a bus indicates whether its placement is in an appropriate cluster. Silhouette coefficient (s_{im}) of bus i can be calculated as

$$s_{im} = \frac{eb_{im} - ac_{ij}}{\max(ac_{ij}, eb_{im})}, \quad m = 1, \ldots, k-1; \quad m \neq j \tag{27}$$

The average of silhouette coefficients, s_{im}, of buses is evaluated as s_i. The average silhouette coefficient, s_i, of all the buses in the whole network is evaluated to give a perspective of average closeness of all the buses to their neighbouring clusters. The coefficient is in a range $[-1, +1]$, where $+1$ indicates that buses are far away from their closest neighbouring clusters; while -1 indicates that the buses are closer to their neighbouring clusters.

4. *Selecting value of k based on silhouette value analysis* Steps 1, 2 and 3 are repeated for different values of k and the one with average silhouette coefficient closest to $+1$ is selected.

□

Flowchart

The proposed AED-based bus clustering method integrates the improved AED-based k-means++ algorithm and silhouette value analysis to group buses for network equivalence of large power system networks. The flowchart shown in Fig. 7 describes this bus clustering method.

Power network equivalencing based on AED-based *k*-means++ clustering method

Power network equivalents based on aggregation of buses in a cluster

The given power system network is divided into various sub-networks governed by local utilities and connected by several tie-lines. Power system operators employ various approximation methods to quickly analyse the behaviour of power system. To replicate the scenario, we employ a similar methodology that divides a given sub-network into smaller areas. This allows us to identify the tie-lines connecting the smaller areas. Further our AED-based improved clustering method is then used to identify appropriate clusters within each smaller area from which we obtain equivalent network. The equivalent network includes each cluster represented by an equivalent bus with a combined generation and load for that cluster connected directly to it. This equivalent bus is connected to the other equivalent buses through the tie-lines. A sample aggregation is illustrated in Fig. 8. In Fig. 8a, a part of the network is shown with two identified clusters connected through a tie-line. Each cluster can be approximated by considering a single aggregated bus without any change in relevant information. As a result, we obtain a simple approximated network with each aggregated bus representing its respective cluster. This is shown in Fig. 8b.

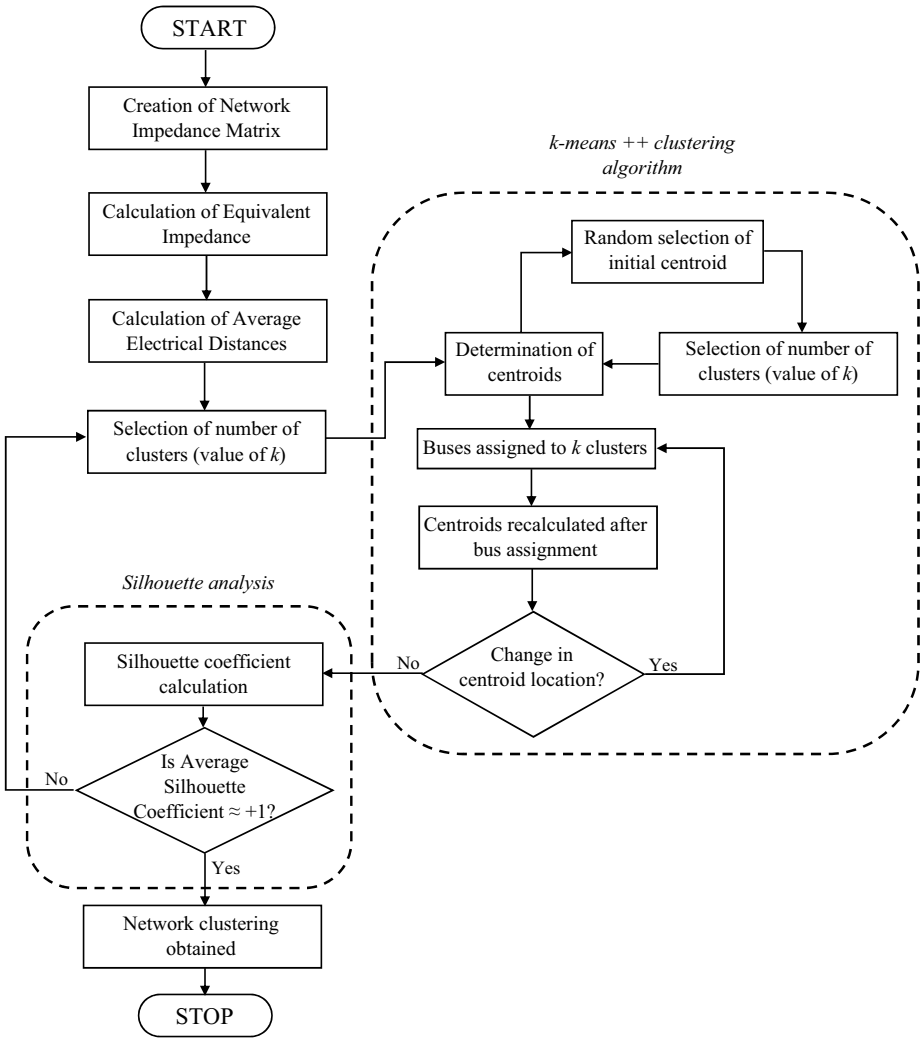


Fig. 7 Flowchart for the proposed improved *k*-means++-based clustering process

Case studies using tie-lines

In this section, we first demonstrate in "39-bus system (Algorithm 3 and method of "AED-based *k*-means bus clustering method)" section the efficacy of the proposed AED-based improved *k*-means++ clustering method by comparing it with our previous clustering method [23] ("AED-based *k*-means bus clustering method" section) on the IEEE 39-bus system. Then in "300-bus system" section, we use IEEE 300 bus system to show the superiority of the proposed method compared to the widely used GSF-based clustering method [5, 6, 21, 22].

39-bus system (Algorithm 3 and method of "AED-based k-means bus clustering method" section)

The IEEE 39-bus system is a standard test system composed of 39 buses with 10 generators and 18 loads connected as shown in Fig. 9 [42, 43]. The net generation and load capacity are 5266.69 MW and 5222.80 MW, respectively. In our study, we consider the

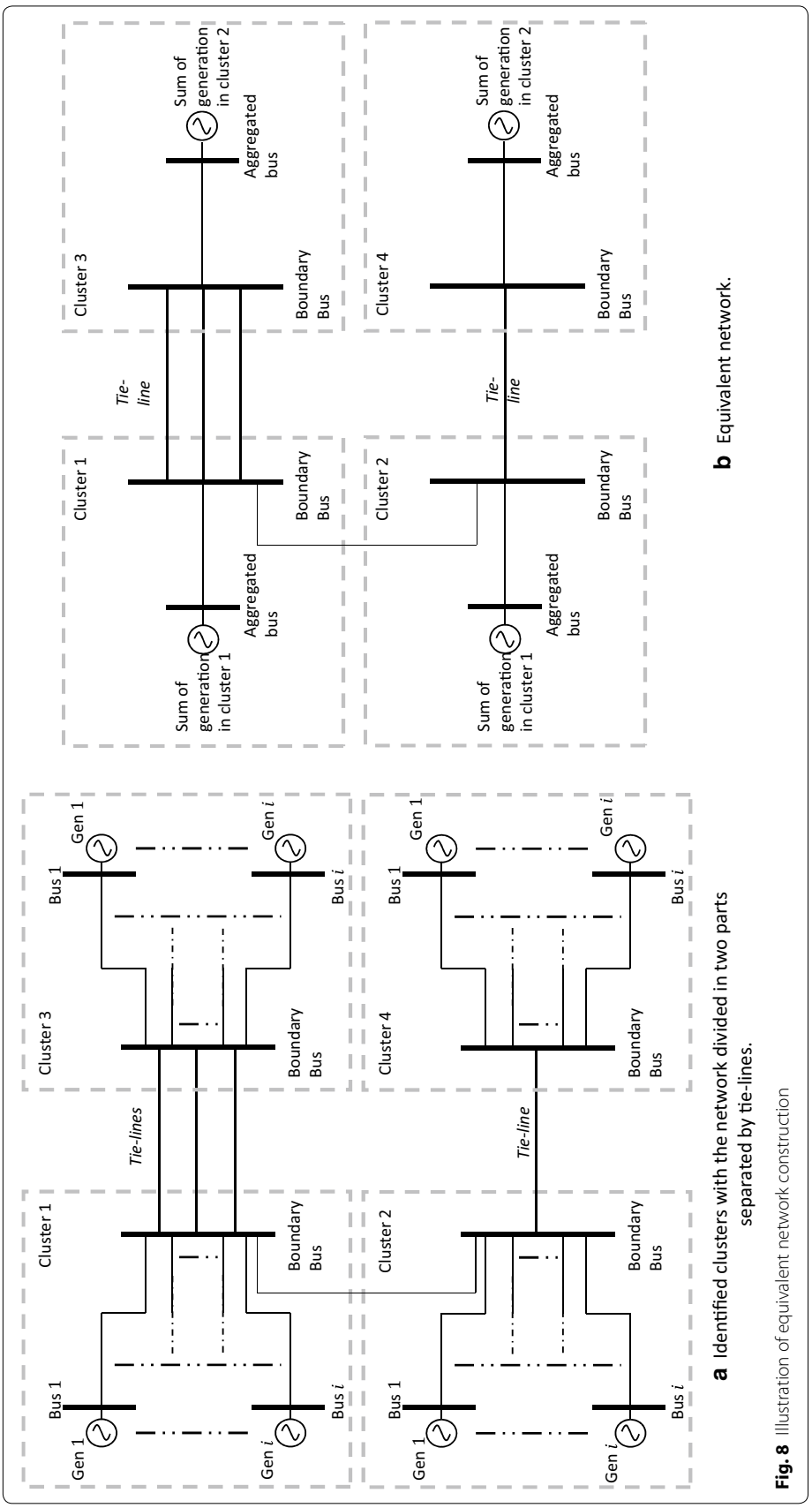


Fig. 8 Illustration of equivalent network construction

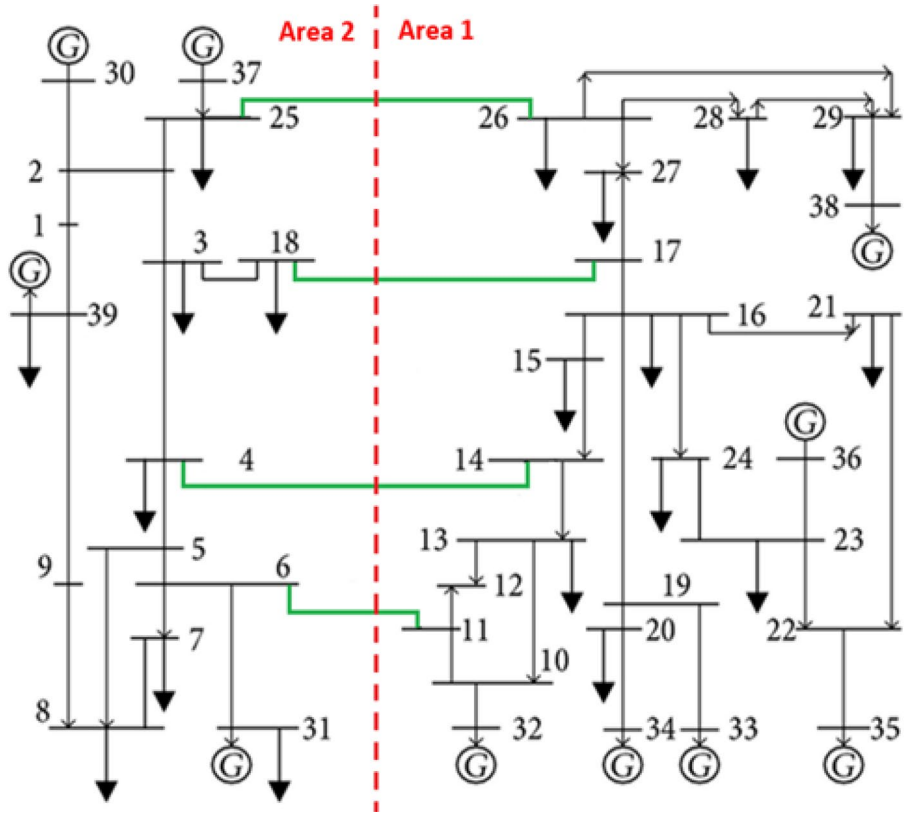


Fig. 9 IEEE 39-bus system

system to be divided into two areas interconnected with four tie-lines in order to analyse the impact of any injection change in generation/load zones on designated transmission line flows. In Fig. 9, area 1 comprises 24 buses, while area 2 consists of 15 buses. Bus 31 is chosen as slack bus.

In our study, the two tie-line scenario is used to compare the proposed method with the method of "AED-based *k*-means bus clustering method" section. In this scenario, any two tie-lines out of four are utilized to connect the two areas in the system. Under this scenario, the clusters are identified using the two methods which are based on average AEDs of each bus with respect to the two tie-lines connected. The two clustering methods follow different algorithms, and hence, the clusters obtained using the two methods are different as observed in Table 1. This may affect the accuracy of tie-line flows in equivalent networks compared to the tie-line flows in the original network. The comparison of accuracy of tie-line flows in the equivalent network is demonstrated using tie-line flow analysis. Further, we also analyse the quality of clusters in the equivalent networks based on the similarity of buses in each cluster.

Tie-line flow analysis (Algorithm 3 and method of "AED-based *k*-means bus clustering method" section) The tie-line flow analysis is based on the comparison of accuracy of net tie-line flows in the equivalent networks with those in the original network. The equivalent networks are created using Algorithm 3 and method of "AED-based *k*-means bus clustering method" section. Different cases with combinations of two tie-lines are studied. In this

Table 1 Comparison of clusters identified by two different AED-based clustering methods

Tie-line 25–26 and Tie-line 17–18			
Clusters identified by AED-based *k*-means algorithm ("AED-based *k*-means bus clustering method" section)		Clusters identified by proposed AED-based improved *k*-means++ algorithm (Algorithm 3)	
Clusters	**Buses**	**Clusters**	**Buses**
SA 11	26, 28, 29, 38	SA 11	26, 28, 29, 38
SA 12	27	SA 12	27
SA 13	10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 32, 33, 34, 35, 36	SA 13	10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 32, 33, 34, 35, 36
SA 21	25, 37	SA 21	25, 37
SA 22	1, 2, 30, 39	SA 22	1, 2, 30
SA 23	18	SA 23	39
SA 24	3, 4, 5, 6, 7, 8, 9, 31	SA 24	3, 18
		SA 25	4, 5, 6, 7, 8, 9, 31

study, the tie-line flows in the original network are calculated using GSFs, and those in the equivalent networks are calculated using average GSFs for each cluster.

Table 2 shows the comparison of net tie-line flows under different tie-line combination cases. It can be observed from Table 2 that the net tie-line flows in the equivalent networks created using the proposed method are more accurate than those in the equivalent networks created using our AED-based *k*-means clustering method of "AED-based *k*-means bus clustering method" section. For instance, in cases 1, 4, 5 and 6, the net flows in the tie-lines in equivalent network obtained using AED-based *k*-means++ method are more accurate compared to the net flows in equivalent network obtained using AED-based *k*-means method. For cases 2 and 3, the net flows in the equivalent networks obtained using the two methods are almost similar. Next, we present a study involving the quality of clusters which explains the tie-line flow analysis on the basis of cluster formations and similarity of buses in each cluster.

Cluster quality analysis We analyse the quality of clusters identified using Algorithm 3 and method of "AED-based *k*-means bus clustering method" section. In this study, the

Table 2 Comparison of net tie-line power flows in the original network and AED-based equivalent networks for 39 bus system

Case no.	Tie-line combination	Original network	AED-based equivalent networks			
		Flow (MW)	AED-based *k*-means algorithm ("AED-based *k*-means bus clustering method" section)		AED-based *k*-means++ algorithm (Algorithm 3)	
			Flow (MW)	% deviation	Flow (MW)	% deviation
Case 1	TL25–26/TL17–18	41.30	41.50	0.48	41.34	0.10
Case 2	TL25–26/TL4–14	35.78	35.81	0.09	35.75	0.09
Case 3	TL25–26/TL6–11	41.50	40.50	2.40	40.49	2.43
Case 4	TL17–18/TL4–14	10.67	19.73	84.90	18.85	76.66
Case 5	TL17–18/TL6–11	23.58	40.52	71.82	32.76	38.93
Case 6	TL4–14/TL6–11	32.81	32.67	0.43	32.77	0.12

cluster quality analysis is carried out in terms of the similarity of buses in clusters. The similarity is measured by the average of standard deviation of the AEDs in each of the clusters identified by the algorithm in [23] and the *k*-means++ algorithm of this paper. The standard deviation of the buses in a cluster can be defined as

$$\sigma_j = \sqrt{\frac{1}{n_j}\sum_{i=1}^{n_j}(d_{ij} - \mu_j)^2}, \tag{28}$$

where σ_j represents the standard deviation of AEDs in jth cluster; d_{ij} indicates the AED of ith bus in jth cluster; and μ_j indicates the mean of AEDs in jth cluster. In (28), a smaller value of standard deviation indicates that the AEDs of the buses tend to be closer to the mean of the cluster, i.e., the buses are closer to each other.

We analyse the cluster quality based on (28), the results of which are shown in Fig. 10. The figure compares the average of standard deviation of the clusters identified by the AED-based improved *k*-means++ clustering method to those identified with our previous method. In Fig. 10, for each case of combinations of tie-lines, grey bars indicate the average of standard deviations of AEDs for the clusters obtained using AED-based improved *k*-means++ algorithm; black bars correspond to the average of standard deviations of AEDs for the clusters obtained using previous AED-based algorithm.

It can be observed from Fig. 10 that for each set of two tie-lines, the similarity of buses in clusters created using proposed AED-based *k*-means++ algorithm is clearly more compared to the buses in clusters created using previous AED-based method. Based on similarity of buses in the clusters, the created equivalent network has net tie-line flows very similar to the original network which can be observed from Table 2. Thus, modifying the clustering method by replacing our method of "AED-based *k*-means bus clustering method" section with the proposed AED-based improved *k*-means++ algorithm provides much accurate results for network clustering schemes.

300-bus system

The IEEE 300-bus system is a test system developed by the IEEE Test Systems Task Force in 1993 [44]. It is composed of 300 buses with 69 generators and 195 loads connected through 409 transmission lines as shown in Fig. 11. Similar to IEEE 39-bus system, in order to analyse the impact on designated transmission line flows, the system is divided into two major areas interconnected using four tie-lines. In Fig. 11, area 1 comprises of 111 buses with total load of 11824.31 MW; area 2 consists of 189 buses and total load of 11,701.54 MW. Bus 7049 is chosen as slack bus.

To demonstrate the efficacy of our Algorithm 3, we compare it with the widely used GSF-based clustering method. We use different two tie-line scenarios in the 300-bus system and the two methods, based on average AEDs and average GSFs respectively, are used to identify clusters to obtain equivalent networks using *k*-means++ algorithm. Further, to validate the use of *k*-means++ algorithm, we compare Algorithm 3 with the AED-based method of "AED-based *k*-means bus clustering method" section that uses *k*-means algorithm. In this study, we use different two tie-line scenarios in

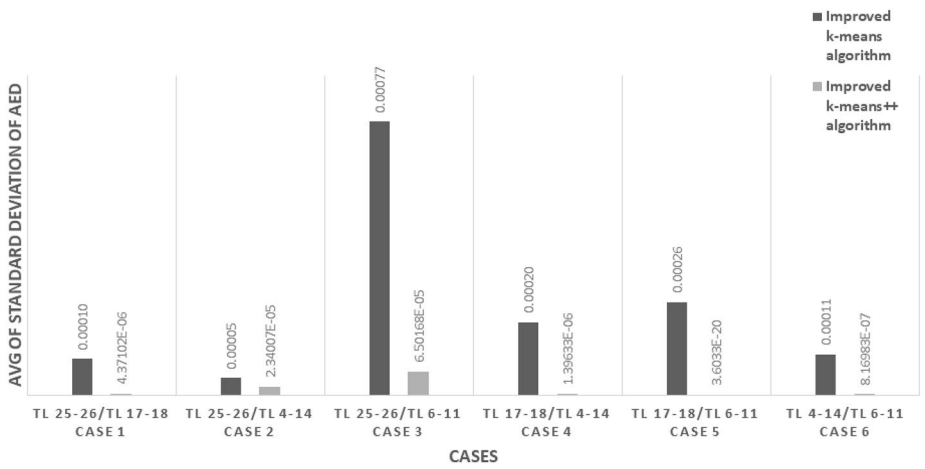


Fig. 10 Comparison of averages of standard deviations of AEDs in clusters for different methods for 39-bus system

the 300-bus system for clustering. These methods yield different clusters for the same network and in order to compare the two clustering methods, we analyse the accuracy of tie-line flows in the equivalent networks. We also analyse the quality of clusters in these equivalent networks in terms of similarity of buses in each cluster.

Tie-line flow analysis (Algorithm 3, method of "AED-based *k*-means bus clustering method" section and GSF-based clustering method) In this section, we compare the net tie-line flows in the equivalent networks with those in the original 300-bus network. In order to study the flows, we consider the cases in which different combinations of two tie-lines connect the areas in the network. The results are shown in Tables 3 and 4 with different combinations of two tie-lines. The net tie-line flows in the equivalent networks in Table 3 are obtained using average GSFs for cluster identification. It can be observed that for each of the cases, the net tie-line flow in the equivalent network created using our proposed method is more accurate as compared to the tie-line flow in equivalent network created using the GSF-based method that uses *k*-means++ algorithm. Also, from Table 4, it can be observed that the tie-line flows in the AED-based equivalent network using *k*-means++ algorithm are better as compared to the equivalent network that uses *k*-means algorithm. Further, cluster quality analysis is done in order to compare the methods based on the similarity of buses in each of the cluster obtained in equivalent networks.

Cluster quality analysis (Algorithm 3 and GSF-based clustering method) In this section, we analyse the quality of clusters identified using the proposed method (Algorithm 3) and GSF-based method in terms of similarity of buses. Here, the similarity of buses refers to the degree of change in AEDs or GSFs in a cluster with respect to a set of tie-lines. The similarity of buses in a cluster is decided using the standard deviation of buses given by (28). Based on (28), we obtain results as shown in Fig. 12. In this figure, for each combination of two tie-lines, grey bars indicate the average of standard deviations of AEDs for the clusters obtained using AED-based improved *k*-means++ algorithm; black bars correspond to the average of standard deviations of GSFs for the clusters obtained using widely used GSF-based clustering method.

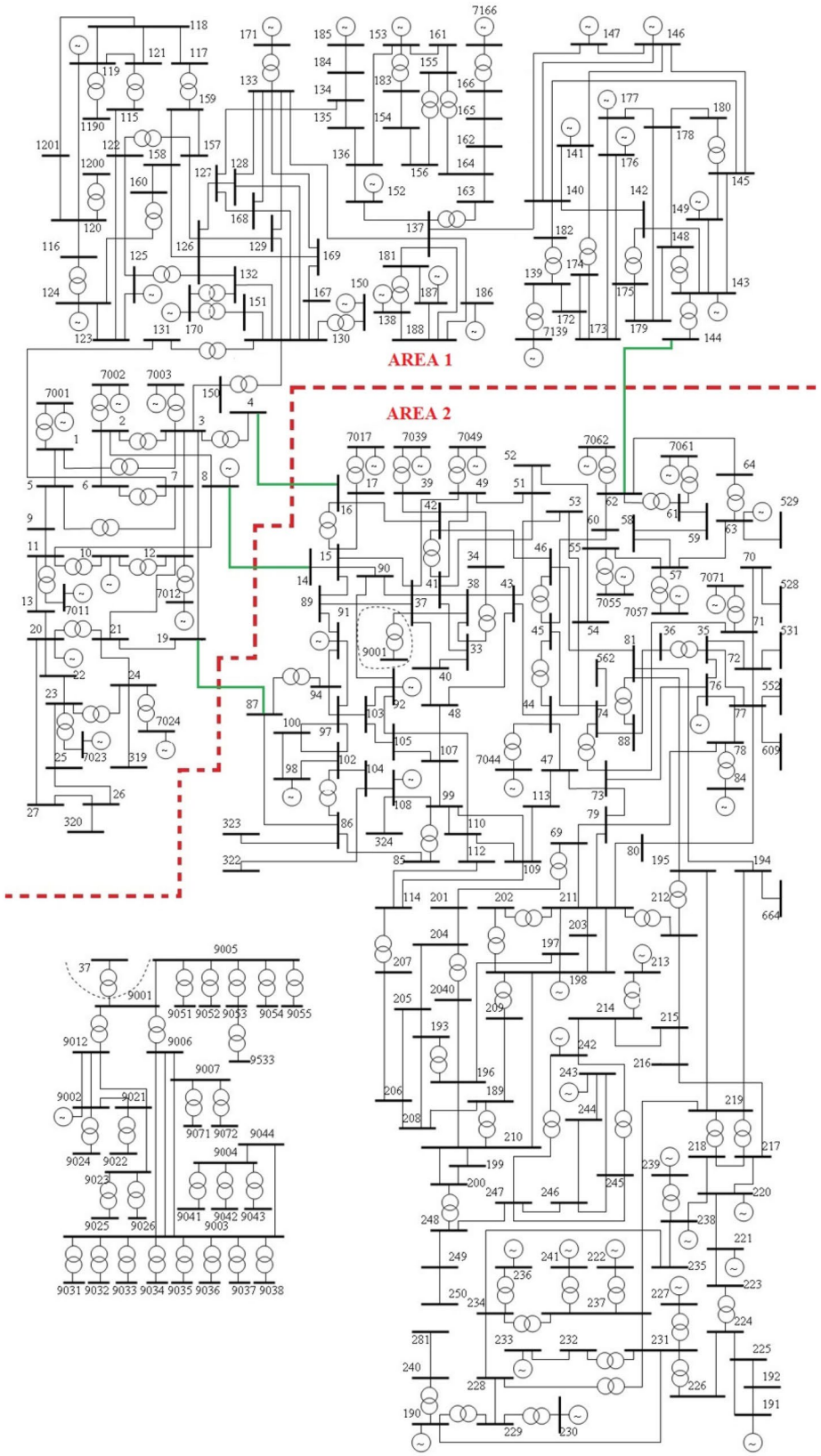


Fig. 11 IEEE 300-bus system

Table 3 Comparison of net tie-line power flows in the original network and those in GSF and AED-based equivalent networks obtained using Algorithm 3 for 300 bus system

Case no.	Tie-line combination	Original network	GSF-based equivalent network		AED-based equivalent network (Algorithm 3)	
		Flow (MW)	Flow (MW)	% deviation	Flow (MW)	% deviation
Case 1	TL19–87/TL8–14	448.09	375.31	16.24	392.85	12.33
Case 2	TL19–87/TL4–16	1051.79	1019.24	3.09	1032.31	1.85
Case 3	TL19–87/TL62–144	55.34	156.84	183.41	57.93	4.68
Case 4	TL8–14/TL62–144	450.86	391.79	13.10	487.01	8.02
Case 5	TL4–16/TL62–144	994.46	490.70	50.66	973.19	2.14

Table 4 Comparison of net tie-line power flows in the original network and the AED-based equivalent networks obtained using the Algorithm 3 and method of "AED-based *k*-means bus clustering method" section

Case no.	Tie-line combination	Original network	AED-based equivalent networks			
		Flow (MW)	AED-based *k*-means algorithm ("AED-based *k*-means bus clustering method" section)		AED-based *k*-means++ algorithm (Algorithm 3)	
			Flow (MW)	% deviation	Flow (MW)	% deviation
Case 1	TL19–87/TL8–14	448.09	292.39	28.08	392.85	12.33
Case 2	TL19–87/TL4–16	1051.79	978.73	6.95	1032.31	1.85
Case 3	TL19–87/TL62–144	55.34	89.87	62.39	57.93	4.68
Case 4	TL8–14/TL62–144	450.86	324.26	28.08	487.01	8.02
Case 5	TL4–16/TL62–144	994.46	922.42	7.24	973.19	2.14

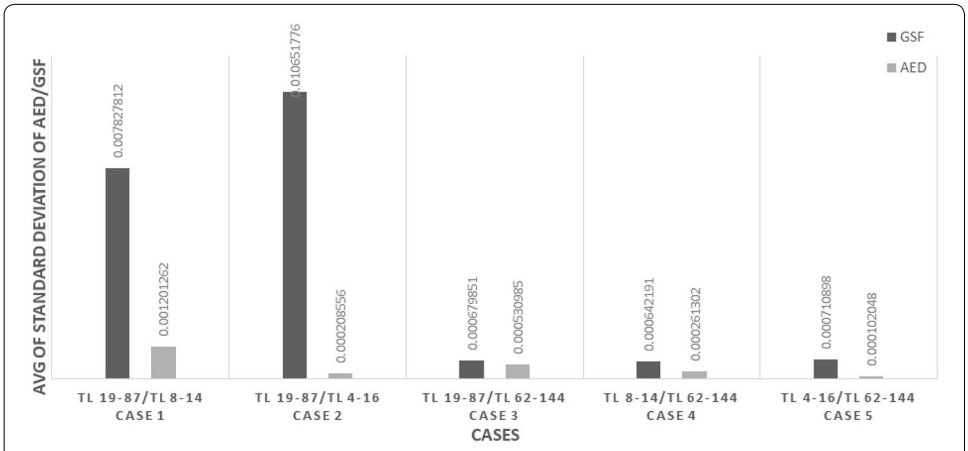


Fig. 12 Comparison of averages of standard deviations of AEDs and GSFs in clusters for 300-bus system

It can be observed from Fig. 12 that for each set of two tie-lines, buses in clusters created based on AEDs are more similar than those in clusters created based on GSFs since the average of standard deviations of AEDs in clusters for each set of two tie-lines is smaller than the average of standard deviations of corresponding GSFs. Based on the

similarity and closeness of buses in the clusters, the created equivalent network has net tie-line flows very similar to that of the original network. It can be observed from Table 3, that the net tie-line flows in the equivalent network of 300-bus system created by bus clustering method using AED-based improved *k*-means++ algorithm are more accurate than those in the equivalent network created by GSF-based bus clustering method. Similar results comparing AED-based *k*-means algorithm [23] and the proposed *k*-means++ algorithm are shown in Table 4.

Conclusion

In this paper, we have presented an AED-based improved bus clustering method for network equivalence of large interconnected power systems. The method utilizes AED-based improved *k*-means++ algorithm for grouping similar buses together to form clusters on the basis of their respective AEDs. The new algorithm is obtained by augmenting the AED-based *k*-means algorithm to probabilistically initialize the centroids of clusters thereby, as in [26], improving the accuracy of the algorithm. The use of silhouette analysis along with improved *k*-means++ algorithm has resulted in further maximizing the accuracy of the clusters. The proposed method has been compared with our previous method [23] on the IEEE 39-bus system. It has been shown that when compared to the full network, the net tie-line flows in the equivalent networks created using the proposed method are more accurate than those in the equivalent networks created using our previous method. Also, the proposed method yields a better cluster quality which shows that the buses in clusters formed using the proposed method are more closely connected than those in the clusters formed using our previous AED-based method.

Moreover, the proposed method has been compared with the widely used GSF-based clustering method [6] on the IEEE 300-bus system. It has been shown that the net tie-line flows in the network with different combinations of tie-lines are more accurate for the equivalent network obtained using the proposed AED-based improved *k*-means++ clustering method than the one obtained using GSF-based clustering method as well as the one obtained with the AED-based *k*-means algorithm. Further, the results of the cluster quality analysis show that the buses in the clusters obtained using proposed method are more closely connected. Thus, the reduced network obtained using the proposed method gives a better representation of the original network compared to the widely used GSF-based clustering method.

In "Implications and relevance for social network analysis" section, we discussed the relevance and implications of our work in the context of social network analysis. We conclude this paper by pointing to an application to what is called the community detection problem in social networks. A community in a social network is a collection of closely related nodes with respect to a closeness measure. A detailed discussion of the community detection problem is given previously [29]. Electrical distance is a measure of closeness of two nodes when the links are assigned weights that capture the characteristics of interest. The Kirchhoff index [45–47] of a cluster is the sum of electrical distances between all pairs of nodes in the cluster. The smaller the Kirchhoff index of a cluster, the closer are the nodes in the cluster. On the other hand, the sum of the AEDs of all the nodes in a cluster is a measure of the total flow across inter-cluster links. Let us call this as inter-cluster Kirchhoff index. Then the smaller

the inter-cluster Kirchhoff index of two clusters, the less closely connected are the nodes in the two clusters. A problem of interest is designing a clustering algorithm that determines clusters such that the Kirchhoff index of each cluster and inter-cluster Kirchhoff index of each pair of clusters are within pre-specified limits. The clusters so determined will provide a solution to the community detection problem in social networks. This is a fairly complex problem involving the solution of a bi-criteria optimization problem, Kirchhoff indices of clusters and inter-cluster indices of pairs of clusters, optimization problem. We are currently studying design of approximation heuristics of this community detection problem. This problem is also related to the problem of partitioning power networks with the aim of containing the impact of cascades due to failures in the network.

Authors' contributions
DS, DW and JNJ conceived of the presented idea of average electrical distance (AED) for network equivalencing. DS and JNJ developed the theory and DS performed the computations related to AED-based network equivalencing method and its comparison with the existing generation shift factor (GSF)-based method. KT, DW and JNJ verified the results and encouraged DS to extend the study by integrating *k*-means++ algorithm and apply it on a bigger network. JNJ and KT supervised the findings of this work. All authors discussed the results and contributed to the final manuscript. All authors read and approved the final manuscript.

Author details
[1] Department of Electrical and Computer Engineering, University of Oklahoma, Norman, USA. [2] Department of Computer Science, University of Oklahoma, Norman, USA. [3] Department of Electrical and Computer Engineering, North Dakota State University, Fargo, USA.

Acknowlegements
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Availability of data and materials
The datasets supporting the conclusions of this article are included within the article and its additional files. The IEEE 39-bus system analysed in this study is commonly known as *"The 10-machine New-England Power System"*. This system's parameters are specified by Athay et al. [42] and are published in a book titled *"Energy Function Analysis for Power System Stability"* [43]. The IEEE 300-bus system test case was developed by the IEEE Test Systems Task Force under the direction of Mike Adibi in 1993. The dataset for this system is available at [44]. Simulation codes related to the study will be available upon request.

Funding
Not applicable.

Publisher's Note

Received: 8 March 2018 Accepted: 3 April 2019
Published online: 22 April 2019

References

1. U.S. Energy information administration: energy explained. http://www.eia.gov/energy_in_brief/article/power_grid.cfm.
2. Wang H, Sanchez CEM, Zimmerman RD, Thomas RJ. On computational issues of market-based optimal power flow. IEEE Trans Power Syst. 2007;22(3):1185–93.
3. Hogan W. A market power model with strategic interaction in electricity networks. Energy J. 1997;18:107–41.
4. Duran H, Arvanitidis N. Simplification for area security analysis: a new look at equivalencing. IEEE Trans Power App Syst. 1972;91(2):670–9.
5. Cheng X, Overbye TJ. PTDF-based power system equivalents. IEEE Trans Power App Syst. 2005;20(4):1868–76.
6. Shi D, Tylavsky DJ. A novel bus-aggregation-based structure preserving power system equivalent. IEEE Trans Power App Syst. 2015;30:4.
7. Dimo P. Nodal analysis of power systems. London: Kent; 1975.
8. Ward JB. Equivalent circuits for power flow studies. AIEE Trans Power App Syst. 1949;68:373–82.
9. Srinivasan S, Sujeer VN, Thulasiraman K. A new equivalence technique in linear graph theory. J Inst Eng. 1964;44(12):496.

10. Srinivasan S, Sujeer VN, Thulasiraman K. Application of equivalence technique in linear graph theory to reduction process in a power system. J Inst Eng. 1966;46:12.
11. Housos EC, Irisarri G, Porter RM, Sasson AM. Steady state network equivalents for power system planning applications. IEEE Trans Power App Syst. 1980;99(6):2113–20.
12. Tinney WF, Bright JM. Adaptive reductions for power flow equivalents. IEEE Trans Power App Syst. 1987;2(2):351–60.
13. Swamy MNS, Thulasiraman K. Graphs, networks and algorithms. New York: Wiley; 1981.
14. Klein DJ, Randic M. Resistance distance. J Math Chem. 1993;12:81–95.
15. Newman MJ. A measure of betweenness centrality based on random walks. Soc Netw. 2005;27(1):39–54.
16. Tizghadam A, Leon-Garcia A. Autonomic traffic engineering for network robustness. IEEE J Select Area Commun. 2010;28(1):39–50.
17. Chellappan V, Sivalingam KM, Krithivasan K. A centrality entropy maximization problem in shortest path routing networks. Comput Netw. 2016;104:1–15.
18. Doyle PG, Snell JL. Random walks and electrical networks. Washington, D.C.: The Mathematical Association of America; 1984.
19. Coppersmith D, Doyle P, Raghavan P, Snir M. Random walks on a 129 weighted graphs and applications to online algorithms. In: 22nd symposium on the theory of computing. 1990. p. 369–78.
20. Blumsack S, et al. Defining power network zones from measures of electrical distance. In: 2009 IEEE power energy society feneral meeting. 2009. p. 1–8.
21. Oh H. A new network reduction methodology for power system planning studies. IEEE Trans Power App Syst. 2010;25(2):677–84.
22. Shi D, Shawhan DL, Li N, Tylavsky DJ. Optimal generation investment planning: pt. 1: network equivalents. In: 44th North American power symposium, Champaign, IL, USA. 2012. p. 1–6.
23. Sharma D, Thulasiraman K, Wu D, Jiang JN. Power network equivalents: a network science-based *k*-means clustering method integrated with silhouette analysis. In: Complex networks and their application VI, Lyon, France. 2017. p. 78–89.
24. Arthur D, Vassilvitskii S. *k*-means++: the advantages of careful seeding. In: 18th annual ACM-SIAM symposium on discrete algorithms. 2007. p. 1027–35.
25. Gutman J, Xiao W. Generalized inverse of the Laplacian matrix and some applications. Bull Acad Serbe Sci Arts. 2004;129(29):15–23.
26. Molitierno JJ. Aplications of combinatorial matrix theory to Laplacian matrices of graphs. Boca Raton: Chapman and Hall–CRC; 2012.
27. Cetinay H, Kuipers FA, Mieghem PV. A topological investigation of power flow. IEEE Syst J. 2018;12(3):2524–32.
28. Brin S, Page L. The anatomy of a large-scale hypertextual web search engine. In: Seventh international world-wide web conference (WWW 1998). 1998.
29. Newman MEJ. Networks: an introduction. Oxford: Oxford Univ. Press; 2010.
30. Merton R. Social theory and social structure. New York: Simon and Schuster; 1968.
31. Lorrain F, White H. Structural equivalence of individuals in social networks. J Math Soc. 1971;1(1):49–80.
32. Rossi RA, Ahmed NK. Role discovery in networks. IEEE Trans Knowl Data Eng. 2015;27(4):1112–31.
33. Faber V. Clustering and the continuous *k*-means algorithm. Los Alamos Sci. 1994;22:138–44.
34. Lloyd SP. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982;28:129–37.
35. Preparata FP, Shamos MI. Computational geometry: an introduction. Berlin: Springer; 1990.
36. Sicotte XB. k-means vs *k*-means++. Cross validated. https://stats.stackexchange.com/q/357606.
37. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc B. 2001;63(2):411–23.
38. Milligan GW, Cooper MC. An examination of procedures for determining the number of clusters in a data set. Psychometrika. 1985;50:159–79.
39. Gordon A. Classification. London: Chapman and Hall–CRC; 1999.
40. Kaufman L, Rousseeuw PJ. Finding groups in data: an introduction to cluster analysis. New York: Wiley; 1990.
41. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20(1):53–65.
42. Athay T, Podmore R, Virmani S. A practical method for the direct analysis of transient stability. IEEE Trans Power App Syst. 1979;2:573–84.
43. Pai MA. Energy function analysis for power system stability. Boston: Kluwer Academic Publishers; 1989.
44. 300 bus Power Flow Test Case Dataset. http://www2.ee.washington.edu/research/pstca/pf300/pg_tca300bus.html.
45. Thulasiraman K, Yadav M. Weighted kirchhoff index of a resistance network and generalization of foster's theorem. In: IEEE International symposium on circuits and systems, Baltimore, USA. 2017. p. 1027–35.
46. Thulasiraman K, Yadav M. Network science meets circuit theory: resistance distance, Kirchhoff index and foster's theorems with generalizations and unifications. IEEE Trans Circuits Syst. 2017;66:1027–35.
47. Yadav M, Thulasiraman K. Network science meets circuit theory: Kirchhoff index of a graph and the power of node-to-datum resistance matrix. In: 2015 IEEE international symposium on circuits and systems (ISCAS). 2015. p. 854–7.