Clustering of Power Networks: An Information-Theoretic Perspective

Mayank Baranwal[1,a] and Srinivasa M. Salapaka[1,b]

***Abstract*— An electrical power transmission system can be represented as a network with nodes and links representing buses and impedance between the buses, respectively. Decomposing a large interconnected power network into smaller loosely-coupled groups facilitates easy and flexible management of the power transmission systems by allowing secondary voltage control at regional levels and controlled islanding that aims to prevent the spreading of large-area blackouts. In this paper, we address the problem of clustering a power network into prespecified number of zones that are weakly coupled *electrically* by using an information-theoretic clustering algorithm. A notion of electrical similarity between electrical buses is developed, where any two buses are considered similar, if reactive perturbations at these buses have a similar effect on the entire network. A thorough theoretical justification of the use of the proposed clustering approach is provided along with the results of our methodology for IEEE test systems.**

I. INTRODUCTION

The North American electrical grid is regarded as the most significant engineering achievement of the 20^{th} century [1], and yet the modern power transmission system faces major challenges due to ever increasing complex interconnections among multiple elements in the grid. Existence of strong links between underlying topological structure and performance in electrical networks have motivated for better strategies for managing and mitigating risks related to network failures. Decomposing a large interconnected power network into smaller loosely-coupled groups facilitates easy and flexible management of the power transmission systems by allowing secondary voltage control at regional levels [2], controlled islanding that aims to prevent the spreading of large-area blackouts, and making the network robust to power and load fluctuations [3].

In this context, it is required to develop interpretable classifications of a given power network. More specifically, the aim of this project is to identify mutually decoupled (or loosely coupled) *clusters* (or zones) in a network such that in any unforeseen event of blackout or catastrophic failure, it is possible to control the spread of power outage and simultaneously identify the nodes that are most affected by the failure. In fact, we propose a clustering metric and approach such that any given node in a network is *tightly* coupled to the nodes within its cluster, while bearing *loose* coupling with nodes in other clusters. Thus the proposed approach reveals the underlying topological structure in the network by decomposing the large network into small number of tractable sub-networks.

[1]University of Illinois at Urbana-Champaign, 61801 IL, USA
[a]baranwa2@illinois.edu, [b]salapaka@illinois.edu
The authors would like to acknowledge NSF grants ECCS 15-09302, CMMI 14-63239 and CNS 15-44635 for supporting this work.

Several recent works have looked at the problem of partitioning of electrical networks using varied approaches. From an abstract viewpoint, an electrical network can be represented by a directed-weighted graph where nodes represent electrical buses in the network, edges representing some notion of electrical connectivity, and weights representing the corresponding *strength* of connectivity. An important element of any graph-clustering approach is the quantification of the notion of *similarity* between any two nodes in a network. These quantifications include but not limited to - (1) *structural similarity*: based on quantities such as degree distribution of nodes and degree assortativity [4], graph diameter [5] and characteristic path length [6], (2) *topological similarity*: Here *electrical* distance is derived either from offline (non realtime) quantities such as nodal conductance matrix [7] or power flow matrix [2] and online quantities, such as derived time-series phase angle data from phase measurement units (PMUs). While the measures of structural similarity are useful for comparing power grids with other graph structures, the absence of any underlying dynamics (arising from Kirchhoff's laws) fails to capture any electrical coupling among nodes of the network. Topological similarity measures alleviate this problem by introducing notion of *electrical distance* obtained using circuit laws and network theorems. Furthermore, offline measures of similarity are preferred since the online methods rely on the observed data after the disturbance has occurred.

In this paper, we quantify the electrical *similarity* between any two nodes based on the first-order perturbation matrix obtained by solving power flow equations [2]. We aim to cluster an electrical network such that nodes within each cluster have *similar* influence over the entire network. The proposed approach is general in the sense that different notions that quantify *similarity* and that quantify *influence* can be used. For ease of exposition, the approach is presented for a particular practical notion of influence; more precisely the influence of one node on another is characterized in terms sensitivity of voltage fluctuations at one node due to reactive power perturbations at the other node. This notion of influence is particularly useful since it encompasses electrical connectivity rather than only the network structure; for instance, two nodes that are strongly electrically coupled through the network even though not directly physically connected to each other will be considered similar in this notion, since voltage variation at one node brings about similar variation at the other node. Note that sensitivity to reactive perturbations strongly assesses the electrical coupling between buses, and not other features such as the amount of power being generated or consumed in the network. The proposed notion of similarity gives a measure of the

electrical coupling between buses for a given reactive power circulation in the network. The approach proposed in this paper is general and can easily accommodate other notions of similarity that depend both on active and reactive powers.

Furthermore, we show that the grouping of nodes (buses) achieved after clustering using this notion of influence is such that the voltage fluctuations at a node due to perturbations at nodes within the same cluster are more than voltage fluctuations due to perturbations at nodes from other clusters. That is, not only that perturbations at two nodes in the same cluster have similar effects on the entire network, the resulting voltage fluctuations at buses from other clusters are much smaller than the voltage fluctuations at the buses from the same cluster. Therefore, the algorithm partitions the electrical network into clusters or zones that are weakly coupled.

In this paper, the problem of grouping buses into electrically similar clusters (or *zones*) is cast as a combinatorial optimization problem, which is reinterpreted as a combinatorial resource allocation problem. Similar combinatorial resource allocation problems have been studied in different areas such as minimum distortion problem in data compression [8], facility location problems [9], graph aggregation [10], motion coordination algorithms,coverage control [11] and mobile sensing network problems [12]. These problems are non convex and computationally complex. It is well known that most of them suffer from poor local minima that riddle the cost surface. A variety of heuristic approaches (such as Lloyd's or *k*-means algorithm) have been proposed to address this difficulty, and they range from repeated optimization with different initialization, and heuristics to good initialization, to heuristic rules for cluster splits and merges. In this work, the approach for clustering buses is based on the deterministic annealing (DA) algorithm [13]. This algorithm offers two important features: (1) ability to avoid many poor local optima and (2) has a relatively faster convergence rate when compared to approaches as simulated annealing or Lloyd's/*k*-means algorithms. The DA algorithm shares connections with the computation of rate-distortion functions in information theory [8], [14], where an effective rate-distortion function parameterized by an annealing variable is formulated and this function is deterministically optimized at successively increased values of the annealing parameter.

A significant contribution of this work is that it also provides a way to characterize the sensitivity (robustness) of representing a large network (graph) by a smaller aggregated graph, with respect to the edge-weight parameters in the original network. This characterization is utilized in defining an edge centrality measure to identify the most important edge in a network, which on removal results in maximum change in the structure of the smaller representative graph. The proposed centrality measure is general in the sense that it applies to any regular directed weighted graph. Furthermore, the sensitivity of graph clustering (or aggregation to a smaller representative graph) to a given edge weight is obtained as a byproduct of the proposed DA approach.

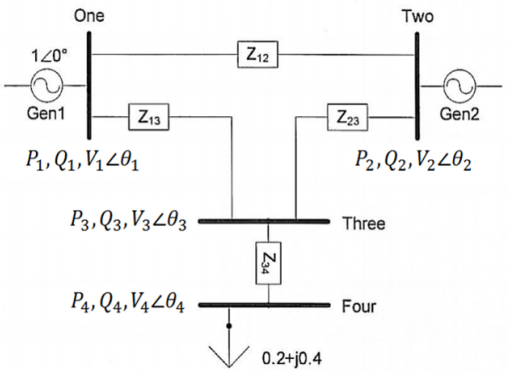

Fig. 1: An example of a four-bus network. Bus 1 is a *Slack* bus whose voltage and angle are specified. Bus 2 is a *Generator* bus (PV bus) whose real output power P and commanded output voltage V are specified. Generators are set to regulate their own bus to a commanded voltage (by adjusting reactive power). Bus 3 is a *connecting* bus with zero load. The load power is known, and this is a PQ bus. Bus 4 is a *load* bus (PQ bus) with per unit (p.u.) active and reactive powers $P = 0.2; Q = 0.4$ both given. The buses are connected by line impedances Z_{ij}. The impedance matrix is denoted by $\mathbf{Z_{bus}}$. The inverse matrix $\mathbf{Y_{bus}} := \mathbf{Z_{bus}}^{-1}$ denotes the admittance matrix.

II. PROBLEM DESCRIPTION

In this paper, we discuss the problem of identifying *similar* buses in an electrical network and grouping them into different zones (clusters). More specifically, we are given a power transmission system with $N \in \mathbb{N}$ buses and electrical connections among them, and the objective is to identify $K \in \mathbb{N}$ zones in the network such that any two buses belonging to the same zone have similar influence over the network. The notion of *electrical distance* and *similarity* is quantified in the next section.

III. QUANTIFICATION OF ELECTRICAL DISTANCE

Our approach for the quantification of electrical proximity of any two nodes is based on computation of the Jacobian matrix obtained by solving power flow equations [2]. In this context, we first describe the fundamental electrical quantities and matrix equations linking them, followed by quantification of electrical distances. Fig. 1 shows a four-bus electrical network, an example network with nodes and links representing buses and corresponding electrical connections, respectively. The buses can be of different types - Slack bus, Generator bus (or PV bus) and Load bus (or PQ bus). Each node i is completely specified by four physical variables - voltage magnitude V_i, phase θ, real power flow P_i, reactive power flow Q_i. The links are specified by line impedances Z_{ij}. $\mathbf{Z_{bus}}$ denotes the impedance matrix of the network. The inverse matrix $\mathbf{Y_{bus}} := [Y_{ij}]$ is the admittance matrix of the network. The current injection at node i is given by I_i. $\mathbf{V}$ and $\mathbf{I}$ are the column vectors of voltage and current magnitudes, respectively. Similarly, $\mathbf{P}, \mathbf{Q}, \mathbf{\Theta}$ are the column vectors depicting real power flows, reactive power flows and the voltage phase angles at the buses of an electrical network.

Recall that the admittance Y_{ij} is generally complex with real part (conductance) G_{ij} and imaginary part (susceptance) B_{ij}, i.e. $Y_{ij} = G_{ij} + jB_{ij}$. These physical variables are related

by the following governing equations

$$\mathbf{I} = \mathbf{Y_{bus}}\mathbf{V}, \qquad \mathbf{V} = \mathbf{Z_{bus}}\mathbf{I},$$
$$P_j = \sum_{k=1}^{N} V_k V_j \left(G_{kj}\cos(\theta_k - \theta_j) + B_{kj}\sin(\theta_k - \theta_j) \right),$$
$$Q_j = \sum_{k=1}^{N} V_k V_j \left(G_{kj}\sin(\theta_k - \theta_j) - B_{kj}\cos(\theta_k - \theta_j) \right), \quad (1)$$

where, N is the number of nodes (buses) in the network and $j \in \{1, \ldots, N\}$. The last two equations are called the power flow equations, and they are necessary to address the *power flow problem*. In a power flow problem, the voltage magnitudes and angles for one set of buses are desired when voltage magnitudes and power levels for another set of buses are known and when a model of the network configuration is available. In order to quantify *electrical distance* between any two nodes of a network, we consider small variations around the power flow solution. The first order perturbations in the above electrical quantities are given by,

$$\mathbf{\Delta I} = \mathbf{Y_{bus}}\mathbf{\Delta V}, \qquad \mathbf{\Delta Q} = [\partial \mathbf{Q}/\partial \mathbf{V}]\mathbf{\Delta V},$$
$$\mathbf{\Delta V} = \mathbf{Z_{bus}}\mathbf{\Delta I}, \qquad \mathbf{\Delta V} = [\partial \mathbf{V}/\partial \mathbf{Q}]\mathbf{\Delta Q}, \quad (2)$$

where matrices $[\partial \mathbf{Q}/\partial \mathbf{V}]$ and $[\partial \mathbf{V}/\partial \mathbf{Q}] \in \mathbb{R}^{N\times N}$ are inverses of each other. While the former matrix appears as a Jacobian during a load-flow computation, the elements of the latter matrix (also known as *sensitivity matrix*) reflect the propagation of voltage variations due to reactive power injection at a node throughout the electrical transmission system.

Note that (2) not only represents the dynamical behavior of an electrical system, it also captures the couplings between different nodes of the grid. Using these equations, it is possible to study the sensitivity of an electrical variable ($\mathbf{V}, \mathbf{I}, \mathbf{P}$, or $\mathbf{Q}$) or any combination of them to perturbations of electrical variables at other nodes. Grouping of nodes based on such sensitivities prove very useful for subsequent resource allocation or power network management problems. For instance, (2) can be used to study the effect of injecting power at a particular node on the voltage magnitudes at the remainder of the network. Alternatively since reactive power management is critical to voltage control for inductive grids, one can study the effect of perturbations of reactive power at a node on the voltages at different nodes in the network. The proposed methodology in this paper can address grouping of nodes based on a combination of sensitivity measures such as above. For ease of illustration, in this paper we investigate the case where we are interested in studying the sensitivity of voltage fluctuations caused at a node with respect to reactive power fluctuations at another node; in particular we consider inductive power networks where the effect of phase-angle perturbations on reactive power at each node is negligible. Most high power electrical networks are indeed largely inductive and therefore exhibit active-reactive decoupling, i.e., $\mathbf{P}$ primarily depends on $\mathbf{\Theta}$ and is almost independent of $\mathbf{V}$ and similarly $\mathbf{Q}$ depends primarily on $\mathbf{V}$ and is independent of $\mathbf{\Theta}$ [15].

The *influence* of one node on another node is given by the magnitude of voltage coupling between the two nodes, which is quantified in terms of *matrix of attenuation* $[\alpha_{ij}] \in \mathbb{R}^{N\times N}$, that is

$$\Delta V_i = \alpha_{ij}\Delta V_j, \qquad \text{where} \quad \alpha_{ij} := \left[\frac{\partial V_i}{\partial Q_j}\right] \Big/ \left[\frac{\partial V_j}{\partial Q_j}\right], \quad (3)$$

which quantifies the voltage fluctuation at node i per unit voltage fluctuation at jth node, when reactive perturbations are applied at node j. Note that the normalization in the definition of α_{ij} has two distinct advantages - (i) making the quantities dimensionless, (ii) assigning equal importance to all the nodes (i.e. $\alpha_{ii} = 1, \forall i$). If $\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_j$ denote the i^{th} and j^{th} columns of the matrix of attenuation, respectively, then the *electrical distance* between nodes i and j is defined as

$$d(i,j) = \|\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_j\|_2^2 = \sum_{k=1}^{N} (\alpha_{ki} - \alpha_{kj})^2; \quad (4)$$

Qualitatively, two nodes i and j are *close*, when the *influence* of these nodes on the network (including the nodes i and j themselves) are commensurate with one another. Note that from the definition (3), the diagonal terms of the attenuation matrix satisfy $\alpha_{kk} = 1$, for all $1 \le k \le N$, and therefore for any $\epsilon > 0$, if $d(i,j) < \epsilon \quad \Rightarrow |\alpha_{ii} - \alpha_{ij}| = |1 - \alpha_{ij}| < \epsilon$.

Similarly, we have $|1 - \alpha_{ji}| < \epsilon$. Therefore $|\alpha_{ij} - \alpha_{ji}| < 2\epsilon$. Therefore, if two nodes i and j are *close*, then as a consequence the influence of perturbations at nodes i and j on each other are *similar*. This observation implies that if we partition the nodes of a network in terms of how *similar* they are in influencing the network, then the influence of nodes on each other from the same cell in a partition will be large, that is close to 1.

With this notion of distance between buses, we view an electrical network as a weighted directed graph (digraph), where buses represent the nodes, the elements α_{ij} represent the edge weights. This makes it amenable to a graph aggregation method developed in [10], where a large weighted directed graph $\mathcal{G}_x$ with N nodes is approximated by a smaller weighted directed graph $\mathcal{G}_y$ (with $K \ll N$ nodes) such that the smaller graph is the *best representation* of the larger graph; the extent of representation is quantified in terms a *dissimilarity* measure. In the resulting smaller graph, each node of $\mathcal{G}_y$ can be viewed as representative of a set of nodes on the larger graph $\mathcal{G}_x$; in fact, the algorithm explicitly gives the set of nodes in $\mathcal{G}_x$ that each node of $\mathcal{G}_y$ represents. Thus this graph aggregation can be used to cluster nodes in $\mathcal{G}_x$ into K clusters, for a given notion of distance between two nodes. Accordingly we use the graph aggregation method to group the buses in the electrical network into clusters for the above notion of electrical distance. In the next section, we briefly present this graph aggregation algorithm and present its important features. A more rigorous and exhaustive treatment can be found in [10]. An important aspect of this article is that we reinterpret this algorithm in terms of a specific information theoretic view point. This reinterpretation enables answering questions such as identifying the most influential edges or couplings in the electrical network; disrupting which can cause the maximum change to the behavior of the electrical network.

IV. GRAPH CLUSTERING

In this section, we develop a framework for aggregating *directed weighted* graphs (digraphs). The objective is to cluster buses in the electrical network; however we present this work for a general directed weighted graph, since this result is important by itself. A digraph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{W})$ is

described in terms of $\mathcal{V}, \mathcal{E} \in \mathcal{V} \times \mathcal{V}$ and $\mathbf{W} \in \mathbb{R}_+^{|\mathcal{V}|\times|\mathcal{V}|}$ which represent the set of *nodes*, *edges* and the *edge-weight matrix*, respectively. Furthermore, $|\mathcal{V}| = N \in \mathbb{N}$ and the relative node weights are denoted by $\{p_i\}, i \in \{1, \ldots, N\}$, which satisfy $p_i \geq 0$ with $\sum_i p_i = 1$. The *incoming vector* of the i^{th} node is described by the weights of its incoming edges and is denoted by $\mathbf{W}_i \triangleq [W_{1i}, \ldots, W_{Ni}]^T$, the ith column of the matrix $\mathbf{W}$. We consider a distance between two nodes i and j based on *edge connectivity* given by $d(\mathbf{W}_i, \mathbf{W}_j)$. Note that this distance measures *similarity* between nodes; for example, small value of $d(\mathbf{W}_i, \mathbf{W}_j)$ implies that nodes i and j have similar connectivity in the graph.

In graph clustering problems, a small representative graph $\mathcal{G}_y$ with $|\mathcal{V}_y| = K$ is obtained from a large graph $\mathcal{G}_x$ with $|\mathcal{V}_x| = N \gg K$ by aggregating *similar* nodes in $\mathcal{V}_x$ into K *supernodes* and then determining the resulting connections among these supernodes. This partition of the nodes $\mathcal{V}_x$ into K clusters, where each cluster is represented by a supernode in $\mathcal{V}_y$ is represented by *partition function* $\phi : \mathcal{V}_x \to \mathcal{V}_y$ which is such that for any $1 \leq j \neq l \leq K$, $\phi^{-1}(j) \subset \mathcal{V}_x$ is non-empty, $\phi^{-1}(j) \cap \phi^{-1}(l) = \emptyset$ and $\cup_{j=1}^{K} \phi^{-1}(j) = \mathcal{V}_x$. Each partition function ϕ defines an *aggregation matrix* $\mathbf{\Phi} \in \{0,1\}^{N\times K}$ as

$$\Phi_{ij} := [\mathbf{\Phi}]_{i,j} = \begin{cases} 1 & \text{if } \phi(i) = j, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Before we state the graph aggregation problem precisely, we present an example for ease of exposition of the subsequent concepts. Consider a graph $\mathcal{G}_x$ with $\mathcal{V}_x = \{1,2,3,4\}$ with $|\mathcal{V}_x| = N = 4$ nodes. The corresponding edge-weight matrix is given by

$$\mathbf{X} = \begin{bmatrix} 0 & 0 & 0 & 0.5 \\ 0 & 0 & 0 & 1.5 \\ 0 & 0 & 0 & 1 \\ 2 & 2 & 2 & 0 \end{bmatrix}.$$

Suppose we want to determine a graph $\mathcal{G}_y$ with two supernodes ($|\mathcal{V}_y| = K = 2$), that is $\mathcal{V}_y = \{1', 2'\}$, which aggregates the graph $\mathcal{G}_x$. In this example, note that $\mathbf{X}$ contains duplicated columns, which indicates $\{1,2,3\}$ are similar; in fact have identical connectivities. Therefore it is easy to see that the a supernode (say $1'$) should correspond to the three nodes $1, 2,$ and 3 and another ($2'$) should correspond to the node 4; that is we have the partition function given by $\phi : \{1,2,3,4\} \to \{1', 2'\}$ with $\phi^{-1}(1') = \{1,2,3\}$ and $\phi^{-1}(2') = \{4\}$. Therefore the corresponding aggregation matrix $\mathbf{\Phi}$ and the weight matrix $\mathbf{Z} \in \mathbb{R}^{N\times K}$ are given by

$$\mathbf{\Phi} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} 0 & 0.5 \\ 0 & 1.5 \\ 0 & 1 \\ 2 & 0 \end{bmatrix}.$$

Note here that the column $\mathbf{Z}_{\phi(i)}$ approximates the ith column of X; in fact, in this example they are exactly the same. The element Z_{kl} in this weight matrix can be interpreted as a directed weight from the kth supernode to lth node. Since the first three rows correspond to the first supernode, this matrix can again be aggregated to obtain the weight matrix $\mathbf{Y}$ of $\mathcal{G}_y$, that is $\mathbf{Y} = \mathbf{\Phi}^T\mathbf{Z} = \begin{bmatrix} 0 & 3 \\ 2 & 0 \end{bmatrix}$, which defines the aggregated graph.

Note that in this example, the graph aggregation essentially required aggregation of of the columns of the matrix $\mathbf{X}$; more precisely, it required finding a partition matrix $\mathbf{\Phi}$ and a corresponding weight matrix $\mathbf{Z}$ such that the cost function $\min_{\mathbf{\Phi},\mathbf{Z}} d(\mathbf{X}_i, \mathbf{Z}_{\phi(i)})$ is minimized. Once the optimal double $(\mathbf{\Phi}, \mathbf{Z})$ are obtained the aggregated graph weight matrix is given by $\mathbf{Y} = \mathbf{\Phi}^T\mathbf{Z}$. Accordingly a general problem of aggregating a large graph $\mathcal{G}_x(\mathcal{V}_x, \mathcal{E}_x, \mathbf{X})$ with $|\mathcal{V}_x| = N$ into a graph $\mathcal{G}_y(\mathcal{V}_y, \mathcal{E}_y, \mathbf{Y})$ with $|\mathcal{V}_y| = K < N$ is given by

$$\min_{\mathbf{\Phi}\in\chi, \mathbf{Z}\in\mathbb{R}^{N\times K}} p_i d(\mathbf{X}_i, \mathbf{Z}_{\phi(i)}), \tag{6}$$

where χ represents the set of all $\{0,1\}^{N\times K}$ aggregation matrices; the edge-weight matrix $\mathbf{Y}$ is then given by $\mathbf{Y} = \mathbf{\Phi}^T\mathbf{Z}$.

Most heuristics that address the optimization problem (6) suffer from the curse of poor local minima and initialization. In the next section, we describe an efficient iterative algorithm that alleviates the issue of poor initialization and has ability to avoid poor local minima.

V. DETERMINING ZONES USING DETERMINISTIC ANNEALING ALGORITHM: AN INFORMATION-THEORETIC VIEWPOINT

Interestingly, the cost function in the aggregation problem (6) is algebraically the same as the cost function that arises in source coding problem from information theory that we now describe. By making this connection, it becomes possible to avail the existing methods for the source coding problem to solve (6).

A source coding problem is described in the context of transmitting information in a parsimonious manner, where a given dataset of *symbols* $\mathbf{X} = \{\mathbf{X}_i\}_{i=1}^N$ is represented by a smaller set of *codewords* $\mathbf{Z} = \{\mathbf{Z}_j\}_{j=1}^K$ with $K \ll N$. The objective is to determine the codewords $\{\mathbf{Z}_j\}_{j=1}^K$ and a corresponding *code* $\phi : \{1, \cdots, N\} \to \{1, \cdots, K\}$ that ascribes to each symbol $\mathbf{X}_i$ a codeword $\mathbf{Z}_{\phi(i)}$ such that a notion of a cumulative representation error given by

$$\min_{\mathbf{Z},\phi} \underbrace{\sum_{i=1}^{N} p_i d(\mathbf{X}_i, \mathbf{Z}_{\phi(i)})}_{D(\mathbf{X},\mathbf{Z})}, \tag{7}$$

where *distortion* $D(\mathbf{X}, \mathbf{Z})$ is a measure of error in representing $\{\mathbf{X}_i\}$ by $\{\mathbf{Z}_{\phi(i)}\}$. Note that this optimization problem is equivalent to the graph aggregation problem in (6). In typical source coding problem formulations, the associations between the symbols and the codewords are not given by hard codes as described above, but are rather ascribed using probabilistic distributions; more precisely the probability of associating a codeword $\mathbf{Z}_j$ to a symbol $\mathbf{X}_i$ is given by $p_{j|i}$, which results in a modified distortion $\bar{D}$ given by

$$\bar{D}(\mathbf{X}, \mathbf{Z}) = \sum_{i=1}^{N} \sum_{j=1}^{K} p_i p_{j|i} d(\mathbf{X}_i, \mathbf{Z}_j). \tag{8}$$

Note that the optimization problem

$$\min_{\{\mathbf{Z}_j\}, \{p_{j|i}\}} \bar{D}(\mathbf{X}, \mathbf{Z}) \tag{9}$$

is equivalent to (7) when we restrict probability distributions $\{p_{j|i}\}$ to be *hard*, that is, $p_{j|i} \in \{0,1\}$ for all $1 \leq i \leq N, 1 \leq j \leq K$.

Since there is no prior information available regarding the choice of $\{p_{j|i}\}$, the deterministic annealing (DA) algorithm seeks $\{p_{j|i}\}$ that minimize the *mutual information* $I(\mathbf{X};\mathbf{Z})$ between $\mathbf{X}$ and $\mathbf{Z}$ without exceeding a given distortion $\bar{D}^*$, i.e., the DA algorithm considers the following problem (also referred to as rate-distortion problem [8])

$$R(\bar{D}) \triangleq \min_{\{p_{j|i}\}} I(\mathbf{X};\mathbf{Z}) \qquad \text{s.t.} \quad \bar{D}(\mathbf{X},\mathbf{Z}) \leq \bar{D}^*, \tag{10}$$

where $I(\mathbf{X};\mathbf{Z}) \triangleq \sum_{i,j} p_i p_{j|i} \log\left(\frac{p_{j|i}}{q_j}\right)$, with $q_j = \sum_{i=1}^{N} p_i p_{j|i}$. (10) appears in the context of rate-distortion theory and provides an analytical expression for how much compression can be achieved using lossy compression methods. The trade-off between minimizing the *mutual information* and minimizing the modified *distortion* is achieved by minimizing the Lagrangian given by, $L \triangleq R(\bar{D}) = I(\mathbf{X};\mathbf{Z}) + \beta\bar{D}(\mathbf{X},\mathbf{Z})$, where β is the Lagrange multiplier, referred to as the *annealing* parameter. Minimizing L with respect to the association $p_{j|i}$ yields a *Gibbs distribution*

$$p_{j|i} = \frac{q_j \exp\{-\beta d(\mathbf{X}_i,\mathbf{Z}_j)\}}{\sum_{j=1}^{K} q_j \exp\{-\beta d(\mathbf{X}_i,\mathbf{Z}_j)\}} \tag{11}$$

By substituting the association weights (11), the Lagrangian simplifies as

$$F(\mathbf{Z}) = -\sum_{i=1}^{N} p_i \log\left(\sum_{j=1}^{K} q_j \exp\{-\beta d(\mathbf{X}_i,\mathbf{Z}_j)\}\right). \tag{12}$$

Note that at very small values of β, minimizing the Lagrangian is equivalent to minimizing the mutual information $I(\mathbf{X};\mathbf{Z})$, which is a convex optimization problem and therefore a closed-form solution can be obtained. Also as β is increased, minimization of the underlying Lagrangian results in the minimization of the modified distortion $\bar{D}(\mathbf{X},\mathbf{Z})$. However, it should be remarked that at large values of the annealing parameter β, the association probabilities $\{p_{j|i}\}$ in (11) are approximately 0 or 1, i.e., $\{p_{j|i}\}$ are hard. Therefore as β increases, from $p_{j|i} \xrightarrow{\beta\to\infty} \Phi_{ij}$, where Φ_{ij} is an element of the (hard) aggregation matrix $\mathbf{\Phi}$, and from the equivalence between (7) and (9) for hard partitions, minimization of (12) with respect to $\{\mathbf{Z}_j\}$ results in minimization of the original distortion function $D(\mathbf{X},\mathbf{Z})$. In the DA algorithm, the Lagrangian in (12) is deterministically optimized at successively increased values of β over repeated iterations (For more details on the DA algorithm see [13]).

VI. IDENTIFYING MOST SIGNIFICANT LINK IN A NETWORK

The solution approach for graph aggregation in this paper has two distinct advantages - (a) The algorithm has ability to avoid poor local minima and is independent of initialization, (b) It provides a way to characterize the sensitivity (robustness) of parameters $\mathbf{Z},\mathbf{\Phi}$ that define the optimal aggregated graph with respect to the edge-weight parameters of the original large graph. This sensitivity calculations can be then used to identify most important edges, which on removal result in significant change in the structure of the smaller representative graph. In this section, we develop a mathematical framework to capture this sensitivity to edge-weight parameters.

Note that the rate function $R(\bar{D})$ provides an analytical expression for how much compression of $\mathbf{X}$ is achieved in representing it by $\mathbf{Z}$, i.e., given a number of codewords $K \in \mathbb{N}$, the central problem in rate-distortion theory is to find the best possible distortion achievable with K codewords. An equivalent problem in the context of graph clustering is to obtain a smaller representative graph containing K supernodes of a given graph, such that the smaller graph approximates the original graph with minimum representation error. Equivalently in the graph aggregation problem, analyzing the sensitivity of the rate-distortion function $R(\bar{D})$ with respect to an element X_{kl} of the edge-weight matrix $\mathbf{X}$ gives a way to quantify the sensitivity of the optimal parameters $\mathbf{\Phi}$ and $\mathbf{Z}$ with respect to X_{kl}.

The sensitivity S_{kl} of representing $\mathbf{X}$ by $\mathbf{Z}$ to an element $X_{kl} \in \mathbf{X}$ is captured by the derivative of the Lagrangian or the rate function $R(\bar{D})$ given by

$$S_{kl} \triangleq \frac{1}{2\beta}\frac{\partial F}{\partial X_{kl}} \xrightarrow{\beta\to\infty} p_l \sum_{j=1}^{K} \Phi_{lj}(X_{kl} - Z_{kj}). \tag{13}$$

From (13) it can be observed that if the node l is associated with the j^{th} cluster, then the sensitivity to an edge emanating from node k and ending at node l is given by $S_{kl} = p_l(X_{kl} - Z_{kj})$. The proposed notion of sensitivity is comprehensive than simply describing the edge with largest weight as the most significant edge in representing a large graph with another smaller dimensional graph. This is seen in the example below.

Consider a graph $\mathcal{G}_x$ with randomly generated edge-weight matrix $\mathbf{X}$ as shown below. $\mathcal{G}_x$ is aggregated into a smaller graph with 5 nodes with $\phi^{-1}(1) = \{1\}$, $\phi^{-1}(2') = \{2,3\}$, $\phi^{-1}(3') = \{4\}$, $\phi^{-1}(4') = \{5\}$ and $\phi^{-1}(5') = \{6\}$.

$$\mathbf{X} = \begin{bmatrix} 5 & 1 & 2 & 0 & 1 & 3 \\ 6 & 8 & 9 & 6 & 10 & 8 \\ 6 & 3 & 9 & 10 & 4 & 3 \\ 4 & 2 & 6 & 1 & 4 & 9 \\ 6 & 10 & 0 & 2 & 6 & 0 \\ 5 & 10 & 3 & 1 & 1 & 7 \end{bmatrix}$$

Interestingly, the proposed sensitivity analysis for the given graph suggests $6 \to 4$ as the most significant edge even though the corresponding edge-weight is just 1. By removing the edge $6 \to 4$ and re-aggregating the graph using DA results in the new partition function given by $\tilde{\phi}^{-1}(1) = \{1,4\}$, $\tilde{\phi}^{-1}(2') = \{2\}$, $\tilde{\phi}^{-1}(3') = \{3\}$, $\tilde{\phi}^{-1}(4') = \{5\}$ and $\tilde{\phi}^{-1}(5') = \{6\}$. Removal of an edge with larger weight does not necessarily result in a different representation. For instance, removing the edge $6 \to 2$ (with an edge-weight of 8) instead does not change the original partition function ϕ when aggregated using DA.

In the context of clustering of power transmission systems, identifying the most significant link corresponds to identifying two nodes which have important bearing on the underlying topology of the network. In the next section, we illustrate the application of DA and the proposed sensitivity measure for some standard test case configurations.

VII. RESULTS AND DISCUSSION

In this section, we report the results from the proposed graph clustering algorithm on some standard network configurations such as the IEEE-14 test case. The IEEE 14 Bus Test Case represents a portion of the American Electric Power System (in the Midwestern US) as of February, 1962. The test case includes all different kinds of buses - Slack, PV and PQ. The matrix of attenuation $\left[\alpha_{ij}\right]$ is first obtained by load-flow computation using Newton-Raphson method. The obtained matrix is then clustered into 3 partitions. Reactive power limits of generators are ignored during load-flow computations for simplicity.

Effect of perturbations on inter and intra-cluster elements: Table I shows the clustering results for the IEEE 14 bus data. The matrix of attenuation is obtained at the operating point (*power flow solution*), and the DA algorithm is employed to obtain three natural partitions (marked by different colors in the 'Bus Type' column and also indicated by the corresponding initials). Column 3 indicates the solution of the power-flow problem in p.u. (per unit), while columns 4 and 5 indicate the effect of perturbations in generator voltage magnitudes at buses 2 and 6, respectively. As is seen in columns 4 and 5 of Table I, the *influence* of these perturbations is larger at the buses belonging to the same group (cluster) where the perturbations originate. For instance, doubling the generator voltage at bus 2 results in change in voltage magnitudes at buses 4 and 5 by about 0.4 p.u. The effect of this perturbation is less severe at other buses, which do not belong to the group formed by the buses $2,3,4$ and 5. Note that bus 3 is a generator bus (PV bus) where voltage is set a priori, and hence there is no change in its voltage magnitude.

Effect of perturbations at buses within the same cluster over the remainder of the network: As stated earlier, two buses are deemed *close* when they have close similarity in terms of the *influences* over the entire network. Columns 4 and 6 in Table I demonstrate the effect of doubling generator voltages at buses 2 and 3, respectively. Note that these buses lie in the same group. It is easily seen that the *influences* of these perturbations over the entire network are *similar*. For instance, bus 12 is not *largely influenced* by these perturbations. The perturbations result in the changes in voltage magnitudes at bus 12 by 0.0126 p.u. and 0.0072 p.u., respectively. However, the effect is *large* on buses such as bus 7, where the changes in voltage magnitudes are 0.1751 p.u. and 0.0993 p.u., respectively.

TABLE I: Clustering Results For IEEE-14 Bus Data

Bus #	Bus Type	Volt mag. Operating pt	Volt mag. ($2\times V_2$)	Volt mag. ($1.5\times V_6$)	Volt mag. ($2\times V_3$)
1	Slack (B)	1.0600	1.0600	1.0600	1.0600
2	PV (Y)	1.0450	2.0900	1.0450	1.0600
3	PV (Y)	1.0100	1.0100	1.0100	2.0200
4	PQ (Y)	1.0142	1.4043	1.1021	1.2372
5	PQ (Y)	1.0172	1.4082	1.1269	1.1495
6	PV (G)	1.0700	1.0700	1.6050	1.0700
7	PQ (B)	1.0503	1.2254	1.1747	1.1496
8	PV (B)	1.0900	1.0900	1.0900	1.0900
9	PQ (G)	1.0337	1.2003	1.2530	1.1270
10	PQ (G)	1.0326	1.1705	1.3088	1.1096
11	PQ (G)	1.0475	1.1179	1.4510	1.0866
12	PQ (G)	1.0535	1.0661	1.5695	1.0607
13	PQ (G)	1.0471	1.0715	1.5421	1.0606
14	PQ (G)	1.0213	1.1276	1.3650	1.0807

Remark: The DA algorithm is scalable and has also been verified for relatively larger networks, such as IEEE-30 bus and IEEE-300 bus systems. The results have not been included for the sake of brevity.

Identifying most significant link in the IEEE 14 bus network: The sensitivity analysis in Sec. VI suggests that the most significant edge is $4 \rightarrow 12$ with $\alpha_{4,12} = 0.8040$. A perturbation in $\alpha_{4,12}$ results in large change in the underlying topological structure of the network. Removal of edge $4 \rightarrow 12$ results in following new partition of the network:

- Cluster 1: Buses $1,2,3,5,7,8$
- Cluster 2: Bus 4
- Cluster 3: Buses $6,9,10,11,12,13,14$

Note that even though $\alpha_{1,4} = 1.0316$ is larger, removal of edge $1 \rightarrow 4$ does not bring about any change in the original partition of the network. Thus a weight based edge centrality measure is inadequate in capturing the true topological structure of an electrical network.

REFERENCES

[1] W. A. Wulf, "Great achievements and grand challenges," *The Bridge*, vol. 30, no. 3, p. 4, 2000.

[2] P. Lagonotte, J. Sabonnadiere, J.-Y. Leost, and J.-P. Paul, "Structural analysis of the electrical system: application to secondary voltage control in france," *IEEE Transactions on Power Systems*, vol. 4, no. 2, pp. 479–486, 1989.

[3] T. Van Cutsem and C. Vournas, *Voltage stability of electric power systems*. Springer Science & Business Media, 1998, vol. 441.

[4] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[5] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[6] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-worldnetworks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[7] E. Hogan, E. Cotilla-Sanchez, M. Halappanavar, S. Wang, P. Mackey, P. Hines, and Z. Huang, "Towards effective clustering techniques for the analysis of electric power grids," in *Proceedings of the 3rd International Workshop on High Performance Computing, Networking and Analytics for the Power Grid*. ACM, 2013, p. 1.

[8] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[9] E. Drezner, "Facility location: A survey of applications and methods," *Journal of the Operational Research Society*, vol. 47, no. 11, pp. 1421–1421, 1996.

[10] Y. Xu, S. M. Salapaka, and C. L. Beck, "Aggregation of graph models and markov chains by deterministic annealing," *Automatic Control, IEEE Transactions on*, vol. 59, no. 10, pp. 2807–2812, 2014.

[11] Y. Xu, S. M. Salapaka, and C. L. Beck, "Clustering and coverage control for systems with acceleration-driven dynamics," *Automatic Control, IEEE Transactions on*, vol. 59, no. 5, pp. 1342–1347, 2014.

[12] J. Cortes, S. Martinez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," in *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, vol. 2. IEEE, 2002, pp. 1327–1332.

[13] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2210–2239, 1998.

[14] K. Rose, "A mapping approach to rate-distortion computation and analysis," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1939–1952, 1994.

[15] G. Andersson, "Modelling and analysis of electric power systems," *EEH-Power Systems Laboratory, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland*, 2004.