Dominant Sets as a Framework for Cluster Ensembles: An Evolutionary Game Theory Approach

Alireza Chakeri, Lawrence O. Hall
Department of Computer Science and Engineering
University of South Florida
Tampa, Florida
Email: chakeri@mail.usf.edu, hall@cse.usf.edu

Abstract—Ensemble clustering aggregates partitions obtained from several individual clustering algorithms. This can improve the accuracy of results from individual methods and provide robustness against variability in the methods applied. Theorems show one can find dominant sets (clusters) very efficiently by using an evolutionary game theoretic approach. Experiments on an MRI data set consisting of about 4 million data are detailed. The distributed dominant set framework generates partitions of quality slightly better than clustering all the data using fuzzy C means.

I. INTRODUCTION

Ensemble Clustering is a relatively new method for clustering with a good description in Strehl and Ghosh [1]. In ensemble clustering, the results of multiple partitions, from one algorithm with different initializations or multiple clustering algorithms applied to the input data, are aggregated into one solution. Ensemble clustering can be used when all attributes of the data set cannot be centrally collected. In this situation, different clustering algorithms can be applied to a restricted number of features or attributes and the results later aggregated. This situation is called feature distributed clustering (FDC) [1]. Also, the ensemble clustering approach can be used when the data (examples) are distributed among several sites. This is called object distributed clustering (ODC), in which each clustering algorithms only accesses a limited number of data points [1].

Recently, several approaches have been applied to the problem of aggregating multiple clustering methods including graph based frameworks [1], voting schemes [2], evidence accumulation clustering (EAC) [3], semidefinite programming [4], and belief function frameworks [5]. Some researchers have focused on high-resolution representations for ensembles of clustering solutions [1], [6]–[8]. These approaches have a size on the order of the number of examples in a partition. In [6], it was determined that existing ensemble algorithms suffer from time complexity problems. Hence, both storage and time costs scale poorly as the number of examples becomes very large.

In this paper, we introduce a game theoretic approach for ensemble clustering based on the notion of dominant sets. A dominant set [9] defines a cluster based on the generalization of maximal cliques in weighted edge graphs. Here the definition of a dominant set is generalized to a weighted vertex graph. To decrease the complexity in both memory and time, we establish theorems to find the dominant sets as well as maximal cliques in a weighted vertex graph very efficiently. The proposed approach for cluster ensembles does not require the feature values of examples to produce the final clusters. Hence, we have a knowledge reuse framework capable of privacy preserving data mining. The method can be applied to both feature distributed clustering and object distributed clustering. Its application in the object distributed case is shown in this paper without loss of generality.

We have modeled the ensemble combination problem as a graph where each vertex is the cluster center from a subset of the data. In this regard, vertices are repeated by their associated weights, i.e. the total number of data belonging to each vertex, forming a graph whose vertices’ size is equal to the number of examples in the original data set. Compactness of the clusters in each subset is also considered in the graph, such that vertices corresponding to the same cluster center are connected together relative to the cluster’s compactness. This measure determines representative cluster centers the examples which belong to them. Finding dominant sets in the resultant graph creates the final partitions of the cluster ensemble problem. We show the dominant sets are very efficiently found by reducing the adjacency matrix size using an evolutionary game theoretic approach.

II. PRELIMINARIES

Background on finding dominant sets, on evolutionary game theory and efficiently finding a Nash equilibrium is given.

A. Quadratic Programming for Maximal Cliques

Taking into account that a cluster is a group of objects having more similarity to one another than objects not in the cluster, some authors [10] argued that a maximal clique in a graph is equivalent to the definition of a cluster. Hence, in this subsection, some background on finding maximal cliques is presented. Let $G = (V, E)$ be an undirected graph, where $V = \{v_1, \ldots, v_n\}$ is the set of vertices and $E \subseteq V \times V$ is the set of edges. A subset $C \subseteq V$ of vertices is called a clique if all of its vertices are connected to each other. A clique is called maximal if it is not a subset of other cliques. The maximal clique with maximum cardinality is called the maximum clique.

Let’s assume that the characteristic vector $x^C$ of dimensionality $n$ for a subset of vertices $C \subseteq V$ with cardinality $|C|$. 

is defined as follows \[11], \[12]

\[
x_i^* = \begin{cases} 
\frac{1}{|C|} & \text{if } v_i \in C \\
0 & \text{otherwise} 
\end{cases} \tag{1}
\]

Consider the following quadratic program called the Motzkin-Straus program \[11]\]

\[
\max_x f(x) = x^T A x \\
\text{subject to } x \in \Delta
\]

where \( \Delta = \{x \in \mathbb{R}^n : x \geq 0, \sum x_i = 1\} \) is the standard simplex, and \( A \) is an \( n \times n \) symmetric adjacency matrix \( A_G = (a_{ij}) \) of graph \( G \) such that \( a_{ij} = 1 \) if \( (v_i, v_j) \in E \) and \( a_{ij} = 0 \) if \( (v_i, v_j) \notin E \). Bomze \[13\] provided a one to one correspondence between local maximizers of \( \Theta \) and the characteristic vectors of the maximal cliques of \( G \) with \( \Theta = A_G + \alpha I \), where \( 0 < \alpha < 1 \) and \( I \) is the identity matrix.

\[\text{B. Quadratic Programming for Dominant Sets}\]

A classical approach to pairwise clustering is to map data points to the nodes of a weighted edge graph with edge weights representing similarity relations among vertices. Let’s consider an undirected weighted edge graph \( G = (V, E, \Theta_{edge}) \), where \( V = \{v_1, \ldots, v_n\} \) is the set of vertices and \( E \subseteq V \times V \) is the set of edges and \( \Theta_{edge} : E \rightarrow [0,1] \) is the edge’s weight function. We represent \( G = (V, E, \Theta_{edge}) \) with its corresponding \( n \times n \) similarity matrix \( A = (a_{ij}) \), where \( a_{ij} = \Theta(v_i, v_j) \). In \[9\], the notion of a Dominant Set was introduced as a definition of a cluster arising from the generalization of maximal cliques in weighted edge graphs. This approach provides a measure of a cluster cohesiveness as well as a degree of belonging of every vertex to each cluster. In \[9\], it was proved that in an unweighted edge graph, dominant sets are equal to strictly maximal cliques.

Pavan et al \[9\] established a one to one correspondence between dominant sets and strict local maximizers of a quadratic form over the standard simplex. If \( D \subseteq V \) is a dominant set, then its characteristic is a strict local solution of program \( \Theta \) with \( A \) being a similarity matrix of the weighted edge graph \( G \). Conversely, if \( x \) is a strict local solution of program \( \Theta \), then its support \( D = \sigma(x) \), where \( \sigma(x) = \{i | x_i > 0\} \), will be a dominant set.

\[\text{C. Game Theoretic Viewpoint of Maximal Clique and Dominant Sets}\]

Recently, a novel connection between finding maximal cliques and dominant sets using quadratic programming and evolutionary game theory was established \[9\], \[13\]–\[16\]. In this regard, a one to one correspondence between asymptotically stable points of the replicator dynamic \[17\] and maximal cliques and dominant sets based on \( \Theta \) was proposed.

In particular, consider problem \( \Theta \) again, where in this case, \( A \) is an \( n \times n \) symmetric payoff matrix. In \[13\], a relation between an Evolutionary Stable Strategy (ESS) of two player games with payoff matrix \( A \) and the quadratic program \( \Theta \) was established, such that \( x \) is the ESS of a two player game with payoff matrix \( A \) if and only if it is a strict local solution of problem \( \Theta \). As a result, in the discrete case, one can find the ESS of the game with payoff matrix \( A = A_G + \alpha I \) to find the maximal cliques of graph \( G \) with adjacency matrix \( A_G \). Similarly, in the continuous case, one can find the ESS of the game with payoff matrix \( A \) to find the dominant sets of the graph with the similarity matrix \( A \).

Now, \( x \in \Delta \) is asymptotically stable under replicator dynamics if and only if \( x \in \Delta \) is a strict local maximizer of problem \( \Theta \), and if and only if \( x \in \Delta \) is an ESS for the equivalent symmetric game \[9\].

\[\text{D. Reduction of Matrix Game to Compute Nash Equilibrium}\]

The Nash equilibrium is a set of strategies such that each player’s strategy is an optimal response to the other players’ strategies \[18\], \[19\]. It was proved that every finite game has at least one Nash equilibrium. In this paper, to find an equilibrium, we use the recursive approach in \[20\] that decomposes a Nash equilibrium computation problem into several smaller problems, such that the solution of the original game can easily be computed from solutions of the smaller games.

\[\text{III. DOMINANT SET FRAMEWORK FOR CLUSTER ENSEMBLES}\]

First, maximal cliques and dominant sets are generalized to a weighted vertex graph. Then, to decrease the complexity, we establish some theorems to find dominant sets very efficiently. Finally, we introduce an approach for ensemble clustering based on generalized dominant sets.

\[\text{A. Maximal Clique in Weighted Vertex Graphs: A Generalization Scheme}\]

Consider an undirected weighted vertex graph. We define the Expanded Graph (EG) of a weighted vertex graph in which weighted vertices repeat the same number as their corresponding weights such that they are fully connected to each other as follows

\[\text{Definition 2: Consider an undirected weighted vertex graph} \ G = (V, E, R_{vertex}), \text{ where } V = \{v_1, \ldots, v_n\} \text{ is the set of vertices, } E \subseteq V \times V \text{ is the edge set, and } R_{vertex} = \{r_1, \ldots, r_n\} \text{ is the set of integer weights of the vertices, i.e. } \forall i, r_i \in \mathbb{N}. \text{ The expanded graph } EG(G) = (V_{EG}, E_{EG}) \text{ is defined as } V_{EG} = \{v_1, v_2, \ldots, v_{1r_1}, \ldots, v_{nr_n}\}, \text{ and } E_{EG} \subseteq V_{EG} \times V_{EG} \text{ such that}\]

\[
\begin{align*}
(v_{ij}, v_{ik}) & \in E_{EG} \quad \forall i \in \{1, \ldots, n\}, \forall j \neq k \in \{1, \ldots, r_i\} \\
(v_{ij}, v_{mn}) & \in E_{EG} \quad \text{if } (v_i, v_m) \in E, \\
& \quad \forall i \neq m \in \{1, \ldots, n\}, \\
& \quad \forall j \in \{1, \ldots, r_i\}, \forall n \in \{1, \ldots, r_m\} 
\end{align*}
\]

(3)

It should be noted that the size of adjacency matrix \( A_{EG} \) of \( EG(G) \) is \( \sum r_i \times \sum r_j \). For simplicity, we use the weighted vertex graph \( G \) and its expansion \( EG(G) \) interchangeably.

\[\text{Theorem 2: Consider a weighted vertex graph} \ G = (V, E, R_{vertex}) \text{ and suppose } C \text{ is a maximal clique in graph } EG(G). \text{ If } v_{ij} \in C, \text{ then } v_{ik} \in C, \forall k \neq j \in \{1, \ldots, r_i\} \]
A reduced adjacency matrix is a maximal clique in graph \( G \) if and only if \( C = \{v_1, \ldots, v_c\} \) is a maximal clique in graph \( G_1 = (V, E) \).

**Proof:** Due to lack of space, proofs can be found in a longer version of this work at: http://www.cse.usf.edu/~hall/dominantfull.pdf

According to Theorem 2, finding a maximal clique in graph \( E(G)(G) \) (and equivalently in graph \( G \)) is the same as finding a maximal clique in the graph \( G_1 = (V, E) \), such that \( G_1 \) is the same as \( G \) without repetition of its vertices, i.e., the weight of each vertex is equal to one.

Next, we tackle the problem of finding local solutions of the quadratic program (2) in the weighted graph \( G \). We define the reduced adjacency matrix of an expanded graph to decrease the size of its adjacency matrix.

**Definition 3:** Consider a weighted vertex graph \( G = (V, E, R_{\text{vertex}}) \) and its expanded graph \( E(G) \). Define the reduced adjacency matrix \( A_R^G \) (\( a_{ij} \) \((a_{ij}) \in [0, 1]\)) of expanded graph \( E(G) \) such that

\[
a_{ij} = \begin{cases} 
1 & \text{if } (v_i, v_j) \in E, i \neq j \\
0 & \text{if } (v_i, v_j) \notin E, i \neq j \\
\frac{\theta_{||v_i|\times|v_j|}}{r_i} & \text{if } i = j 
\end{cases} \quad (4)
\]

As \( r_i \) increases, \( a_{ii} \) approaches 1. As it approaches one, it is more likely to be in a bigger clique (i.e., it is repeated more).

**Definition 4:** Consider a weighted vertex graph \( G = (V, E, R_{\text{vertex}}) \), where \( V = \{v_1, \ldots, v_n\} \), and its expanded graph \( E(G) \). Assume that \( C \) is a maximal clique in \( E(G) \), then the reduced characteristic vector \( x_R^C \) with size \( n \) of a maximal clique \( C \) as follows

\[
(x_R^C)_i = \begin{cases} 
\frac{\theta_{||v_i|\times|v_j|}}{r_i} & \text{if } v_{ij} \in C, \forall j \in \{1, \ldots, r_i\} \\
0 & \text{otherwise}
\end{cases} \quad (5)
\]

Based on definitions 3 and 4, we establish the following theorem to find local solutions of a weighted vertex graph very efficiently in terms of solving the quadratic program (2) with the reduced adjacency matrix of its expanded graph.

**Theorem 3:** Consider a weighted vertex graph \( G = (V, E, R_{\text{vertex}}) \) and the reduced adjacency matrix \( A_R^G \) of \( E(G) \). Then the local solutions of quadratic program (2) with the matrix \( A = A_R^G \) are in the form of reduced characteristic vectors \( x_R^C \) of maximal cliques \( C \) of \( E(G) \).

**Property 1:** Since the reduced adjacency matrix \( A_R^G \) (\( a_{ij} \in [0, 1] \)) of the expanded graph \( E(G) \) is symmetric, the ESSs of the corresponding game with payoff matrix \( A_R^G \) can be found using replicator dynamics.

Consequently, by Theorem 3 and Property 1, instead of finding a maximal clique in a graph with weighted vertices using a quadratic equation having an adjacency matrix with size \( \sum_{v_i} r_i \times \sum_{v_i} r_i \), or equivalently finding ESSs of the corresponding game with payoff matrix with size \( \sum_{v_i} r_i \times \sum_{v_i} r_i \), we can find the maximal cliques efficiently only by solving the quadratic program with its reduced adjacency matrix of size \( n \times n \) using replicator dynamics.

---

**Example 1:** Consider the graph \( G = (V, E, R_{\text{vertex}}) \) in which \( V = \{v_1, v_2, v_3, v_4\} \) and its expanded graph \( E(G) \) are shown in Fig. 1, where \( G_1 = (V, E) \) is the same graph with repetition of its vertices and \( V_{EG} = \{v_1, v_2, v_3, v_4\} \). Also the adjacency matrix \( A_{EG} \) and the reduced adjacency matrix \( A_2^{R_{EG}} \) of expanded graph \( E(G) \) are shown in Table I.

In addition, the ESSs of the expanded graph \( E(G) \) and its reduced game with adjacency matrix \( A_2^{R_{EG}} \) and \( A_2^{R_{EG}} \), respectively, are shown in Table II which agree with expectations. \( ESS_1 \) and \( ESS_2 \) are the characteristic vectors of maximal cliques \( C_1 = \{v_1, v_3\} \) and \( C_2 = \{v_1, v_2\} \) in the graph \( G_2 \), respectively, i.e., \( ESS_1' \) and \( ESS_2' \) are the reduced characteristic vector of maximal cliques \( C_1 \) and \( C_2 \).

**TABLE I. ADJACENCY MATRIX AND REDUCED MATRIX OF THE EXPANDED GRAPH IN FIG. 1**

<table>
<thead>
<tr>
<th>Adjacency Matrix ( A_{EG} )</th>
<th>Reduced Adjacency Matrix ( A_2^{R_{EG}} )</th>
</tr>
</thead>
</table>
| \[
\begin{bmatrix}
0 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 1 & 1 & 0 \\
0.2 & 1 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}
| \[
\begin{bmatrix}
0.2 & 1 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 1 & 0 \\
0.2 & 1 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}
| 4 \times 4
|}

**TABLE II. ESSS OF THE ORIGINAL AND REDUCED EXPANDED GRAPH**

<table>
<thead>
<tr>
<th>Using Adjacency Matrix ( A_{EG} )</th>
<th>Using Adjacency Matrix ( A_2^{R_{EG}} )</th>
</tr>
</thead>
</table>
| \[
\begin{bmatrix}
\frac{1}{4} & 0 & 0 & 0 & 0 \\
0 & \frac{1}{4} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{4} & 0 & 0 \\
0 & 0 & 0 & \frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{4}
\end{bmatrix}
| \[
\begin{bmatrix}
\frac{1}{4} & 0 & 0 & 0 & 0 \\
0 & \frac{1}{4} & 0 & 0 & 0 \\
0 & 0 & \frac{1}{4} & 0 & 0 \\
0 & 0 & 0 & \frac{1}{4} & 0 \\
0 & 0 & 0 & 0 & \frac{1}{4}
\end{bmatrix}
| 4 \times 4
|}

**B. Dominant Sets in Weighted Vertex and Edge Graphs: A Generalization Scheme**

In this subsection, the dominant sets framework [9] is generalized to a weighted vertex graph similar to the proposed generalization of maximal cliques.

**Definition 5:** Consider an undirected weighted vertex and edge graph \( G = (V, E, \Theta_{\text{edge}}, R_{\text{vertex}}, \beta_{\text{vertex}}) \), where \( V = \{v_1, \ldots, v_n\} \) is the vertex set. \( E \subseteq V \times V \) is the edge set, \( \Theta_{\text{edge}} : E \rightarrow [0, 1] \) is the edge’s weight function, \( R_{\text{vertex}} = \{r_1, \ldots, r_n\} \) is the set of integer weights of vertices, and \( \beta_{\text{vertex}} = \{\beta_1, \ldots, \beta_n\} \) is the set of arbitrary values assigned to each vertex, where \( \beta_i \in [0, 1], \forall i \in \{1, \ldots, n\} \). The expanded graph \( E(G) \) is defined...
with \( V_{\text{EG}} = \{v_{11}, v_{12}, \ldots, v_{m1}, v_{m2}, \ldots, v_{mn}\} \) and \( E_{\text{EG}} \subseteq V_{\text{EG}} \times V_{\text{EG}} \) such that

\[
\begin{align*}
\Theta_{\text{edge}}(v_{ij}, v_{ik}) &= 0 & \forall i \in \{1, \ldots, n\}, \forall j \in \{1, \ldots, r_i\} \\
\Theta_{\text{edge}}(v_{ij}, v_{ik}) &= \beta_i & \forall i \in \{1, \ldots, n\}, \forall j \neq k \in \{1, \ldots, r_i\} \\
\Theta_{\text{edge}}(v_{ij}, v_{mn}) &= \Theta_{\text{edge}}(v_{ij}, v_{nm}) & \forall i \neq m \in \{1, \ldots, n\}, \forall j \in \{1, \ldots, r_i\}, \forall n \in \{1, \ldots, r_m\}
\end{align*}
\]

As is customary, we represent \( G = (V, E, \Theta_{\text{edge}}, R_{\text{vertex}}) \) with its corresponding symmetric similarity matrix \( A_{\text{EG}} = (a_{ij,mn}) \) of size \( \sum_{i=1}^{n} r_i \times \sum_{i=1}^{n} r_i \), where \( a_{ij,mn} = \Theta_{\text{edge}}(v_{ij}, v_{mn}) \). Here \( \beta_i \) is a measure of compactness for each cluster.

**Theorem 4:** Let \( G = (V, E, \Theta_{\text{edge}}, R_{\text{vertex}}, \beta_{\text{vertex}}) \) be a weighted vertex and edge graph, and \( D \) be a dominant set in the expanded graph \( E(G) \). If \( v_{ij} \) belongs to the dominant set \( D \), then \( v_{ijk}, v_{ik} \neq j \in \{1, \ldots, r_i\} \), also belongs to \( D \). In addition the degrees of belonging \( v_{ij} \) and \( v_{ik} \) to \( D \) are the same, i.e. \( w_{D}(v_{ij}) = w_{D}(v_{ik}) \), where \( w_{D}(v_{ij}) \) is the weight of the vertex with regard to \( D \) [9].

**Definition 6:** Consider a weighted vertex and edge graph \( G = (V, E, \Theta_{\text{edge}}, R_{\text{vertex}}, \beta_{\text{vertex}}) \) and its expanded graph \( E(G) \). Define the reduced similarity matrix \( A_{R} = (a_{ij}) \) of the expanded graph \( E(G) \) such that

\[
a_{ij} = \begin{cases} 
\Theta_{\text{edge}}(v_{ij}, v_{ik}) & \text{if } i \neq j \\
\beta_i (r - 1) & \text{if } i = j
\end{cases}
\]

As \( a_{ij} \) approaches 1, again the vertex is repeated more and likely to be in a larger dominant set.

**Definition 7:** Consider a weighted vertex and edge graph \( G = (V, E, \Theta_{\text{edge}}, R_{\text{vertex}}, \beta_{\text{vertex}}) \) and its expanded graph \( E(G) \). Assume that \( D \subseteq V_{\text{EG}} \) is a dominant set of \( E(G) \). Then we define the reduced characteristic vector \( x_{R} \) with size \( n \) as follows

\[
(x_{R})_{i} = \begin{cases} 
\frac{w_{D}(v_{ij})}{W(D)} & \text{if } v_{ij} \in D, \forall j \in \{1, \ldots, r_i\} \\
0 & \text{otherwise}
\end{cases}
\]

where \( W(D) \) is the total weight of the vertices in \( D \).

**Theorem 5:** Consider a weighted vertex and edge graph \( G = (V, E, \Theta_{\text{edge}}, R_{\text{vertex}}, \beta_{\text{vertex}}) \) and its expanded graph \( E(G) \). If \( D \subseteq V_{\text{EG}} \) is a dominant set of \( E(G) \), then its reduced characteristic vector \( x_{R}^{D} \), is a strict local solution of program (2) with \( A_{R} \) as the similarity matrix. Conversely, if \( x \) is a strict local solution of program (2) with \( A_{R} \) as the similarity matrix, then based on its support \( D = \sigma (x) = \{v_{m1}, \ldots, v_{mk}\} \), \( D = \{v_{m1}, v_{m2}, \ldots, v_{mr}, \ldots, v_{nk}\} \) will be a dominant set of \( E(G) \).

According to the above theorem, finding a dominant set in a weighted and edge graph \( G = (V, E, \Theta_{\text{edge}}, R_{\text{vertex}}, \beta_{\text{vertex}}) \) is equivalent to finding a dominant set in the graph with the reduced similarity matrix. Consequently, instead of finding a dominant set in a graph with weighted vertices and edges using a quadratic equation having a similarity matrix of size \( \sum_{vi} r_i \times \sum_{vi} r_i \), or equivalently finding ESSs of the corresponding game with payoff matrix of size \( \sum_{vi} r_i \times \sum_{vi} r_i \), we can find the dominant sets efficiently only by solving the quadratic program with its reduced similarity matrix of size \( n \times n \).

---

**C. Dominant Sets as a Cluster Ensemble**

We introduce a new approach for ensemble clustering based on dominant sets. The dominant sets of the modeled graph are considered to generate an ensemble clustering or partition. In other words, the cluster centers are clustered using a dominant sets framework to obtain the final partitions for the cluster ensemble problem. Let us consider \( P = \{p_1, p_2, \ldots, p_l\} \) as a set of \( l \) subsets of some data set, and \( C_i = \{c_{i1}, c_{i2}, \ldots, c_{ik}\} \) as the set of \( k \) cluster centroids in partition \( p_i \). In addition, assume \( W_i = \{w_{11}, w_{22}, \ldots, w_{ik}\} \) is the set of cluster centers’ weights which is computed as the number of examples in partition \( p_i \) which belong to it, and \( \Gamma_i = \{\gamma_{11}, \gamma_{12}, \ldots, \gamma_{ik}\} \) is a set of measure of the compactness of each cluster in partition \( p_i \). Here, we assume that each partition is clustered by the Fuzzy C Means algorithm with same number of clusters \( k \). These assumptions, i.e. the same number of clusters and Fuzzy C Means, are not essential. In the proposed approach, we map the cluster centers to a weighted vertex and edge graph \( G = (V, E, \Theta_{\text{edge}}, R_{\text{vertex}}, \beta_{\text{vertex}}) \), where \( V = \bigcup_{i=1}^{l} C_i = \{v_{11}, \ldots, v_{1xk}\} \). \( E \subseteq V \times V \) is the edge set, \( \Theta_{\text{edge}} : E \rightarrow [0, 1] \) is the edge’s weight function such that \( \Theta(v_{ij}, v_{ij}) = e^{-d(v_{ij}, v_{ij})} \) in which \( d(v_{ij}, v_{ij}) \) is the Euclidean distance between \( v_{ij} \) and \( v_{ij} \), \( R_{\text{vertex}} = \bigcup_{i=1}^{l} W_i = \{r_{11}, \ldots, r_{1xk}\} \) is the set of cluster centers’ weights which is computed as the number of examples in \( p_i \) which belong to it, and \( \beta_{\text{vertex}} = \bigcup_{i=1}^{l} \Gamma_i = \{\beta_1, \beta_2, \ldots, \beta_{1xk}\} \) is a set of measures of the compactness of each cluster.

According to the definition of the expanded graph \( E(G) \), the number of vertices in \( E(G) \) is the same as the number of examples in the data set. The effect of the parameter \( \beta \) is shown in Fig 2. Let’s assume that we have the two clusters (+ and o) obtained from two chunks of data (clearly different distributions, but we will ignore that issue for this example). One can easily see that although the cluster centers (*) are close to each other and might be in the same dominant set, the data points in each cluster clearly do not fit together. The compactness measure will prevent these cluster centers from being in the same dominant set. Based on the proposed framework, one can find the dominant sets of the corresponding expanded graph \( E(G) \) by solving (2) with similarity matrix \( A_{EG} \) of size \( \sum_{vi} r_i \times \sum_{vi} r_i \). However, in large data sets clustering, each \( r_i \), and consequently \( \sum_{vi} r_i \), can become a very large problem since each partition \( p_i \) has many data to be clustered. Theorem 5, allows an efficient solution to this problem by very significantly reducing the size of the similarity matrix. We alternatively can solve the quadratic equation (2) with
TABLE III. CLUSTER CENTERS OF MRI DATA SETS APPLYING FCM AT ONCE

<table>
<thead>
<tr>
<th>Data set</th>
<th>MRI3-16</th>
<th>MRI3-17</th>
<th>MRI3-18</th>
</tr>
</thead>
<tbody>
<tr>
<td>cluster 1</td>
<td>123.1130, 382.6959, 136.2344</td>
<td>230.4890, 552.6390, 144.8076</td>
<td>213.4844, 343.7840, 108.0734</td>
</tr>
<tr>
<td>cluster 2</td>
<td>256.8882, 386.1485, 438.9061</td>
<td>382.5229, 482.9379, 370.5186</td>
<td>390.0435, 604.3740, 485.9799</td>
</tr>
<tr>
<td>cluster 3</td>
<td>175.0433, 740.2540, 282.3662</td>
<td>275.6511, 784.2197, 297.1899</td>
<td>273.5858, 784.3137, 298.2820</td>
</tr>
</tbody>
</table>

TABLE IV. CLUSTER CENTERS OF MRI-13 DATA SET DIVIDED INTO 10, 100 AND 1000 EQUAL Sized CHUNKS

<table>
<thead>
<tr>
<th>Cluster centers</th>
<th>1%</th>
<th>1%</th>
<th>0.1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRI3-16</td>
<td>(221.0099, 380.5132, 136.5247)</td>
<td>(222.1126, 383.0009, 136.6873)</td>
<td>(222.4870, 381.2843, 135.1310)</td>
</tr>
<tr>
<td>MRI3-17</td>
<td>385.0709, 482.2205, 438.7132</td>
<td>396.8853, 483.9198, 433.9557</td>
<td>395.0115, 483.9238, 433.7370</td>
</tr>
</tbody>
</table>

TABLE V. CLUSTER CENTERS OF MRI-17 DATA SET DIVIDED INTO 10, 100 AND 1000 EQUAL Sized CHUNKS

<table>
<thead>
<tr>
<th>Cluster centers</th>
<th>1%</th>
<th>1%</th>
<th>0.1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRI3-16</td>
<td>(220.1297, 381.9471, 136.4687)</td>
<td>(220.0884, 382.8072, 136.9260)</td>
<td>(220.1459, 382.8437, 139.9047)</td>
</tr>
<tr>
<td>MRI3-17</td>
<td>(382.7835, 482.9485, 438.4586)</td>
<td>(382.7835, 483.9386, 433.9895)</td>
<td>(382.7835, 483.9286, 433.5281)</td>
</tr>
<tr>
<td>MRI3-18</td>
<td>(175.6215, 740.1221, 297.4250)</td>
<td>(175.6187, 740.3155, 297.6199)</td>
<td>(175.6165, 740.4031, 297.6991)</td>
</tr>
</tbody>
</table>

TABLE VI. CLUSTER CENTERS OF MRI-18 DATA SET DIVIDED INTO 10, 100 AND 1000 EQUAL Sized CHUNKS

<table>
<thead>
<tr>
<th>Cluster centers</th>
<th>1%</th>
<th>1%</th>
<th>0.1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRI3-16</td>
<td>(221.6427, 382.9200, 136.2270)</td>
<td>(221.3598, 382.0137, 136.9470)</td>
<td>(221.4997, 384.0979, 138.8904)</td>
</tr>
<tr>
<td>MRI3-17</td>
<td>(395.5959, 484.2300, 438.1905)</td>
<td>(395.5625, 484.2036, 438.1985)</td>
<td>(395.5625, 484.1947, 438.1865)</td>
</tr>
<tr>
<td>MRI3-18</td>
<td>(175.6210, 740.4230, 298.2516)</td>
<td>(175.6725, 740.0307, 298.1946)</td>
<td>(175.6840, 740.3175, 298.1522)</td>
</tr>
</tbody>
</table>

TABLE VII. SPEEDUP FOR DIFFERENT SAMPLE SIZE OF MRI DATA SETS

<table>
<thead>
<tr>
<th>Sample size (%)</th>
<th>10%</th>
<th>1%</th>
<th>0.1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRI3-16</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MRI3-17</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>MRI3-18</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

IV. EXPERIMENTS

Large Image Data Sets: MRI-13, 17, 18 are 3-sequence magnetic resonance (MR) image data sets, including proton density (PD), T1 and T2 images of the brain in the axial plane. The data sets consist of an MRI volume with 96 slices for each of the 3-sequence (512×512) 12-bit images. The images are first processed to remove air and skull pixels. The remaining pixels in each of the 96 slice 3-sequence images are unrolled into 3-D (one dimension per sequence) feature vectors. These 3-D vectors from each of the 96 slices are combined, creating a data set comprised of about 4 million (3882771) 3-D objects.

The cluster centers from applying FCM once on each MRI data set is shown in the Table III.

In the following experiments, data was randomized and divided into 10, 100 and 1000 equal sized chunks, i.e. sample sizes in each chunk are 10%, 1% and 0.1% of the total number of examples in the data set. Then, each chunk was clustered using FCM with 3 clusters. Each experiment was repeated 10 times, and each time a new randomization of the data was done. Also, the cluster centers were initialized randomly in each chunk.

In the case of 10 chunks, there are 30 cluster centers after performing FCM on each chunk. Each cluster center vs is also associated with two values ri and γi, such that ri is the number of examples that the cluster i and γi was calculated as

\[ \gamma_i = e^{-\text{avgdist}_i} \]  

where avgdist_i is the average Euclidean distance between the cluster center vi and the samples that belong to the cluster i. In the proposed framework, the dominant sets of the resultant graph, derived from Definition 5, having the number of vertices equal to the total number of examples in the data set, i.e. 3882771, are interpreted as the solution to the cluster ensemble problem. However, based on our proposed theorems, one can reduce the size of the adjacency matrix from 3882771×3882771 to 30×30. From the derived reduced characteristic vectors of the dominant sets in the reduced graph, one can obtain the characteristic vectors of the original dominant sets in the expanded graph according to Theorems 4 and 5. A similar explanation holds for the 100 and 1000 chunks cases. In these cases, the reduced adjacency matrices have the size of 300×300 and 3000×3000, respectively, instead of 3882771×3882771.

In this paper, the approach in [22] was used to enumerate the dominant sets in the reduced graph. Then, for each resultant dominant set, a dominant set center was introduced by calculating the average of the clusters centers which belong to that dominant set. Also, to obtain a specific number of final cluster centers for the whole data set, we used a hierarchical clustering method to aggregate the dominant sets’ centers. First the two closest dominant sets centers were merged to reduce the number of dominant sets by one and the same procedure was followed until the desired number of final sets’ centers, i.e. cluster centers, were obtained.

The cluster centers obtained via the proposed approach on
each data set for three different chunks are shown in Table IV, V and VI. As one can see, the difference between cluster centers applying FCM at once on the whole data set and the proposed approach on different number of chunks are negligible.

Also, the speedup as another evaluation criteria is defined as $\frac{\text{Time of Execution}}{\text{Time of Execution}_{\text{chunked}}}$ - The speedup of different sample sizes are depicted in Table VII. If the data need to be broken into a large number of chunks, e.g. greater than 100, then the feasibility of producing a solution with an extremely large dataset would be more important than speedup.

To show the algorithm performance, the reformulated optimization criteria $R_m$ [23] where the aim is to minimize it, was computed as follows

$$R_m = \sum_{k=1}^{n} \left( \sum_{i=1}^{c} d_{ik} (x_k, v_i) \left( \frac{1}{m} \right) \right) (1-m)$$

(10)

where $x_k$ is the data point, $v_i$ is the cluster center, $d_{ik} (x_k, v_i)$ can be any distance measure between the $k$-th data point and $i$-th cluster center and $m$ is any number greater than 1. In this paper, we used the Euclidean distance and $m = 2$. We compared the quality of the proposed approach by computing the difference between the mean $R_m$ value in 10 experiments using the dominant set approach and clustering all the data at once (global clustering), and then expressing it as a percentage. For example if we denote the mean $R_m$ value of our approach as $m_1$ and the mean $R_m$ value of global FCM clustering as $m_2$, then the difference in quality (DQ) is

$$DQ = \left( \frac{m_1 - m_2}{m_2} \right) \times 100$$

(11)

The results for the MRI3-16 data set are shown in Table VIII. As the Table VIII shows, our proposed approach is even better than clustering all data at once in terms of obtaining a lower mean $R_m$ value. However, the most significant aspect of using this approach is the mathematically efficient way to reduce the complexity of the ensemble clustering problem.

V. CONCLUSIONS

In this paper, we introduced a new approach for the cluster ensemble problem using dominant sets. In this regard, the ensemble clustering problem is transformed into a graph clustering problem. The modeled graph is then searched to find all dominant sets which are interpreted as the final partitions. Also, established theorems based on evolutionary game theory show how to find dominant sets very efficiently in the modeled weighted vertex and edge graph. The experiments on a real image MRI data set show that final partitions can be improved by the proposed approach using subsets of data.

REFERENCES