Assessment of Hamming Distance and Self Organization Map in Solving Cell Formation Problem

Sanaa A Hamza and Ammar J Khalaf

1, 2 Kerbala Technical Institute, Al-Furat Al-Awsat Technical University, 56001 Kerbala, Iraq.
1sanaalihamza62@gmail.com, 1inkr.sna@atu.edu.iq, Phone: 07802143995, 2inkr.emd2@atu.edu.iq

Abstract

The cell formation (CF) problem is considered the most essential issue in cellular manufacturing systems (CMS). CF deals with the arrangement of similar parts into groups known as part families (PFs) and organizes machines also into groups, called machine cells (MCs). In the literature, numerous methods, models and algorithms have been proposed and developed to handle CF problems. However, very few studies have dealt with the assessment and comparison of these methods, to identify the most effective. This has provided strong motivation for the study presented here. The present paper focuses on two methods that are used infrequently to form MCs and PFs, and applies them in three strategies: the first is based on the use of a hamming distance only, while the second uses only a self-organization map (SOM). However, the third method applies a hybrid approach based on SOM and hamming distance. The outputs of the selected methods were compared, to select the best one. A set of five benchmark datasets and three performance measures was used for comparison and evaluation. These performance measures are: percent of the exceptional elements (PE), grouping efficiency (GE), and machine utilization (MU). The results refer to the outperforms of the hamming distance in terms of PE, GE and MU for most of the selected benchmark problems.

Keywords: Cell formation, cellular manufacturing system, grouping efficiency, hamming distance, machine utilization, self-organization map.

1. Introduction

The cellular manufacturing system (CMS) is seen as one of the best strategies to deal with customer requirements and the problem of continuous changing of product designs. CM operates on the basis of group technology thought and has positive impact in terms of quality and productivity.

Cell formation (CF) is the most significant part of the CMS. It deals with the gathering of similar parts into groups called families, and dissimilar machines similarly gathering into groups called cells. In the literature several methods have been proposed for solving the CF problem, such as similarity coefficient-based methods, array clustering-based methods, mathematical programming-based methods, meta-heuristic based methods, etc.

Su et al., [1] proposed an effectual initialization pattern to build an initial SOM map. Then, to make small consequent adjustments and to enhance the accuracy of the initial map, they utilized the self-organizing feature map algorithm. Some data sets were tested to demonstrate the performance of the suggested method.

Venkumar and Haq [2] have used a Kohonen Self-Organizing Map (KSOM) to solve CF problems. The effectiveness of the proposed method was identified by the number of voids and exceptional elements. The proposed method was applied to some benchmarked datasets presented in the open literature. GE was utilized as a performance parameter to compare the results of the suggested method with the best-known results in the literature. The outcome was found to be better than or equal to, when compared with other algorithms, in terms of reducing the number of exceptional elements.

Chattopadhyay et al., [3] applied SOM for solving a CF problem with the aim of increasing the group technology efficiency (GTE). They analyzed the suggested method by utilizing some
benchmark datasets selected from the published research work. Finally, they proved the effectiveness of their approach.

Chattopadhyay et al., [4] utilized two approaches for solving CF problems. The first was known as principle component analysis (PCA) and was used for data extraction. The second was known as SOM and utilized in shaping visual clustering. These two approaches were used for (0,1) incidence matrix with the sequence of operations.

Once more, Chattopadhyay et al., [5] suggested a new method for CF based on the use of SOM. Their method applied to binary (part-machine) matrix where some datasets were chosen from the published research to apply the proposed approach. The results revealed enhancement of the grouping efficacy parameter for 70% of the selected problems.

Hamza and Adesta, [6] introduced a brief review of the literature on the integration of the fundamental decisions on the design of CM systems. These important decisions are: (i) cell formation (CF); (ii) cell layout (CL) and (iii) cell scheduling (CS). The limitations and objectives of the combination of these basic decisions have been recognized. Guidelines for future studies were suggested, including the consideration of another essential issue in CM, namely feasibility assessment (FA). To make the CMS more realistic, the continual consideration of a real life data such as the capacity of machines, sequence of processes, demand for products and alternative process plan has been recommended for inclusion in the integration issue.

Furthermore, Chattopadhyay et al., [7] applied SOM for handling the CF problem, afterward for the large size problems they used SOM in a hierarchical style known as a growing hierarchical self-organizing map (GHSOM). Then, they compared the two proposed algorithms by applying them with fifteen datasets from the open literature and verified an improvement of the grouping efficacy (GC) and GTE for 70% of the data sets.

Potočnik et al., [8] introduced a new method based on SOM for facility layout. The proposed method was divided into two steps, in the first step, SOM was utilized to create PFs and MCs, while in the second step, the layout of the facility was completed in a real-life factory with the consideration of a method for material handling.

Pradhan and Mishra [9] proposed a method based on Minkowski distance to cluster MCs, and SOM to cluster PFs. Throughout the clustering procedure, some bottleneck machines were created, this problem was overcome by manual rearrangement. The outcome of the Minkowski distance and SOM proved the efficiency of the proposed method, with GC equivalent to 96.4%.

Wu and Suzuki [10] developed a new technique for handling the problem of CF. The proposed methodology included two steps. In the first step, a new similarity coefficient (SC) was developed, that involves the number of repeated operations and sequences of operations to create PFs. However, a new mathematical model was applied in the second step. This model contained some features such as operation time, machine capacity, alternative routing, lot splitting and part demand to assign machines to families of parts with minimal operation cost, machine cost and inter cell movement cost. This method compared the cost of duplication of machines and the inter-cell movement cost. The effectiveness of the proposed method was revealed by the testing of some data sets.

Kiriş and Tüysüz [11] identified a good classification of the various CM methods and carried out a performance comparison of different clustering methods for manufacturing.

Hamza [12] assessed the existing production information, and then integrated the results of this phase with the results of CF to attain an effective CMS. In the evaluation phase, some hierarchical procedures were applied while in the design (CF) part, one of the well-known array-based clustering
methods was utilized. This method is called ROC and used to form MCs & PFs. However, in Hamza’s study some significant parameters were applied to estimate the performance of the proposed CM, specifically: grouping efficiency (GE), grouping efficacy (GC), voids (V), exceptional elements (EE), percent of exceptional elements (PE) and machine utilization (MU). To validate this work, three data sets (matrices) were selected. The approach that followed obtained a powerful CM solution.

Tran et al., [13] enhanced a hamming distance by use of Bloom filters based on fuzzy join in MapReduce. Then, the algorithms were evaluated and compared analytically with a cost model.

Hamza and jehad [14] applied a heuristic method based on the hamming distance to create PFs and MCs, their suggested method first computed the hamming distance for the parts, then reorganized them based on the results to form the PFs. After that, the hamming distance was computed for machines, then the machines reordered based on the results to shape the MCs. To validate the proposed method, three datasets from the open literature were used. For evaluation and comparison, five performance measures were utilized, these measures were percent of exceptional elements (PE), grouping efficiency (GE), voids, machine utilization (MU), and exceptional elements (EE). The results referred to the superior performance of the hamming distance-based method compared with the best-known results of the published researches work in the literature.

According to the literature, very few researchers have dealt with the assessment and comparison of CF methods. This has provided strong motivation for the study presented here. The present study focused on the CF issue, whereby two different methods were used to create MCs and PFs. These two methods are known as hamming distance and SOM, and are applied individually and simultaneously. Afterward, these two methods were evaluated via several performance parameters and some benchmark datasets from the literature.

2. Self-Organization Map

A self-organizing map (SOM) is a type of unsupervised learning algorithm based on artificial neural networks. It tries to build a two-dimensional map of a problem space. The self-organizing map uses competitive learning rather than error-correction learning, which constitutes the main difference between a self-organizing map and other neural networks. A visual representation of data on a hexagonal or rectangular lattice can be produced by a self-organizing map. Applications include meteorology, oceanography, project prioritization, oil and gas exploration.

The SOM’s functionality is based on competitive learning; a competition between the output neurons of the network and themselves for activation. The result of this competition is one output neuron, the winner neuron, will go on at any time. This competition enacted by arranging negative feedback between the output neurons, a lattice of one or two dimensions is used: neurons are placed on its nodes and during the competitive learning process the neurons become selectively tuned to the patterns at the input stage. As a result, a certain order between the winner neurons will be established with respect to each other, leading to the creation of a meaningful coordinate system over the lattice for the various input features. Figure 1 shows the schematic diagram of a two-dimensional lattice of neurons that is commonly used [15].

First a synaptic weight is given initially in the network by the SOM formation algorithm. A random-number generator is used to give them small value weights; in such, there will be no prior order on the feature map. In the formation of the self-organizing map there are three essential processes involved once the network has been properly initialized, as summarized here:

*Competition:* The neurons in the network compute their particular values of a distinguished function. For each input pattern, this distinguished function provides the basis for competition between all the
neurons. A winner of this competition is declared the particular neuron with the largest value of distinguished function.

**Cooperation:** To provide the basis for cooperation between neighboring neurons. The locative site of a topological neighborhood of activated neurons is determined by the winning neuron.

**Synaptic Adaptation:** The activated neurons are enabled, by this mechanism, to increase their values individually of the distinguished function using appropriate adjustments applied to their synaptic weights in relation to the input pattern. The winning neuron response is enhanced by the subsequent application of a similar input pattern, such that the adjustments are made [16].

![Figure 1. Two-dimensional lattice of neurons, illustrated for a three-dimensional input and four-by-four dimensional output [15].](image)

3. **Hamming Distance**

Hamming distance is a function that defines a distance between each pair of binary vectors. This distance function is greater than or equal to zero and it is metric, it shows the dissimilarity between the two binary vectors. For fixed length N, the Hamming distance between two binary vectors U and V is (XOR). Giving the number of ones which show how much U is dissimilar to V, it is widely used in information theory to find errors in transmitting codes and it is used in error correction and error section for data transmission. The hamming distance between the two vectors:

\[ d(U; V) = \sum_{i=1}^{N} (U_i \oplus V_i) \]

Is the number of points in which the two vectors are dissimilar, in part-machine cell formation the hamming distance was used between each part/machine to find the least dissimilar parts/machines after treating them as binary vectors [17, 18]. Table 1 refers to the logical relationship for XOR.

| INPUT | OUTPUT |
|-------|--------|
| U     | V      | U XOR V |

Table 1. The logical relationship for XOR.
4. Methodology

The strategy that follows is divided into three sections: in the first section, the hamming distance used for creating MCs and PFs. SOM is applied in the second section to shape MCs and PFs. However, a hybrid method based on SOM and hamming distance was used to create MCs and PFs, and that is described in the third section. Then, three well-known performance measures were used to evaluate the results of the mentioned three methods. These performance measures are PE, GE, and MU.

One dataset was selected from the literature to explain the proposed method. This data set contains eight machines and ten parts (8*10). The proposed methodology (Hamming distance and SOM) was applied successfully by MATLAB R2016b (9.1.0.441655) September 7, 2016 64 bit. Figure 2 refers to the methodology steps. Further details of the proposed method are explained in the following sections.

4.1 First method (H-H)

\[
\begin{array}{ccc}
0 & 0 & 0 \\
0 & 1 & 1 \\
1 & 0 & 1 \\
1 & 1 & 0 \\
\end{array}
\]

Figure 2. The methodology steps.
In this method, the hamming distance was used to build part families firstly, and then to form machine cells. The steps of this method are applied to the matrix (8*10) as in the following. Figure 3 refers to these steps.

1. Calculate the hamming distance for the parts in the selected matrix (8*10) (Figure 3b).
2. Rearrange the matrix according to the difference percentages, the smaller the difference percentage the more parts similarity (Figure 3c).
3. Now apply the hamming distance again for the new arranged matrix, but for the machines this time (Figure 3d).
4. Rearrange the matrix above, according to the new machine difference percentage like before (Figure 3e).
5. The arrangement in step four is called part families and machine cells (Figure 3f).
4.2 Second method (S-S)

In this method, SOM was used to build part families first, then it was used for machines to form the machine cells. The procedures of this method are displayed on the same matrix (8*10) in the following. Figure 4 refers to these steps.

1. Run SOM using MATLAB for the parts in matrix (8*10) (Figure 4b).
2. Rearrange the matrix according to the SOM results based on the similarities in the color where all the parts similar in colors, arrange sequentially and the less similar follow them (Figure 4c).
3. Again, apply SOM for the machines on the matrix (Figure 4d).
4. Rearrange the matrix visually based on the SOM Figures to get the MCs and PFs (Figure 4e).
5. The arrangement in step four is called part families and machine cells (Figure 4f).

4.3 Third method (S-H)

In this method, a hybrid technique was used, where the SOM was used firstly to build part families, then the hamming distance was used to create MCs. The details of this method are displayed as follows. Figure 5 refers to these steps.

1. Run SOM using MATLAB for the parts in matrix (8*10) (Figure 5b).
2. Rearrange the matrix according to the SOM results based on the similarities in the color where all the parts are similar in colors, arrange sequentially and the less similar follow them, (Figure 5c).
3. Apply the hamming distance to the machines in the new arranged matrix (Figure 5d).
4. Rearrange the matrix above, according to the new machine difference percentage (Figure 5e).
5. The arrangement in step four is called part families and machine cells (Figure 5f).
| P1 | P2 | P3 | P4 | P5 | P6 | P7 | P8 | P9 | P10 |
|----|----|----|----|----|----|----|----|----|-----|
| M1 | 1  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0   |
| M2 | 0  | 1  | 1  | 1  | 0  | 0  | 1  | 1  | 0   |
| M3 | 1  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 0   |
| M4 | 1  | 0  | 0  | 0  | 0  | 1  | 0  | 0  | 0   |
| M5 | 0  | 1  | 0  | 1  | 0  | 0  | 0  | 1  | 0   |
| M6 | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 0  | 1   |
| M7 | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 0  | 1   |
| M8 | 0  | 0  | 0  | 0  | 1  | 1  | 0  | 0  | 1   |

(a): The original matrix (8*10)

| P2 | P4 | P8 | P10 | P3 | P7 | P9 | P5 | P6 | P1 |
|----|----|----|-----|----|----|----|----|----|----|
| M1 | 1  | 1  | 0  | 0  | 1  | 0  | 0  | 0  | 0   |
| M2 | 1  | 1  | 1  | 1  | 1  | 0  | 0  | 0  | 0   |
| M3 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0   |
| M4 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1   |
| M5 | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0   |
| M6 | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 1  | 0   |
| M7 | 0  | 0  | 0  | 1  | 0  | 1  | 1  | 0  | 0   |
| M8 | 0  | 0  | 0  | 1  | 0  | 1  | 1  | 1  | 0   |

(c): Rearranging the matrix according to the SOM results based on the similarities in the color

| P2 | P4 | P8 | P10 | P3 | P7 | P9 | P5 | P6 | P1 |
|----|----|----|-----|----|----|----|----|----|----|
| M2 | 1  | 1  | 1  | 1  | 1  | 0  | 0  | 0  | 0   |
| M5 | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0   |
| M1 | 1  | 1  | 0  | 0  | 1  | 0  | 0  | 0  | 1   |
| M6 | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 1   |
| M8 | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 1  | 1   |
| M3 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1   |
| M4 | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 1   |
| M7 | 0  | 0  | 0  | 1  | 0  | 1  | 0  | 0  | 0   |

(e): Rearranging the matrix visually based on the SOM figures

| P2 | P4 | P8 | P10 | P3 | P7 | P9 | P5 | P6 | P1 |
|----|----|----|-----|----|----|----|----|----|----|
| M2 | 1  | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0   |
| M5 | 1  | 1  | 1  | 0  | 0  | 0  | 0  | 0  | 0   |
| M1 | 1  | 1  | 0  | 1  | 0  | 0  | 0  | 0  | 1   |
| M6 | 0  | 0  | 0  | 0  | 1  | 1  | 1  | 1  | 0   |
| M8 | 0  | 0  | 0  | 1  | 0  | 0  | 1  | 1  | 1   |
| M3 | 0  | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1   |
| M4 | 0  | 0  | 0  | 0  | 0  | 0  | 1  | 1  | 1   |
| M7 | 0  | 0  | 0  | 1  | 0  | 1  | 0  | 0  | 0   |

(f): Identifying part families and machine cells

Figure 4. Steps for applying the second method (SOM based method).
Figure 5. Steps for applying the third method (SOM and Hamming distance-based method).
5. **Performance measures**

To validate the quality of the solution, five well-known measures were used to assess the performance of the three proposed methods. These measures are known as voids, EE, PE, GE and MU. These performance measures are explained in the following section.

5.1 **Number of Voids (V)**

Voids refer to the number of zero’s entries in the final created cells, these zeros signify that some parts have no need to operate on some machines, or some machines have idle times and don’t use all of the available capacity.

5.2 **Exceptional Elements (EE)**

The off-diagonal positive entries (1s) which are called the exceptional elements (EE) in the final CF solution can be used to measure the performance of the selected CF method. EE can be computed as in Equation 1:

\[
EE = eo
\]  

(1)

Where, eo: is the number of EEs or the off-diagonal positive entries.

5.3 **Percentage of the Exceptional Elements (PE)**

The clustering quality can be computed by the number of ones that stay outside the diagonal block [19, 20]. Equation 2 represents the PE formula [21,22]:

\[
PE = \frac{EE}{UE} \times 100
\]  

(2)

Where, EE is the numbers of 1s that remain outside the block diagonal, UE refers to the overall number of 1s inside the incidence matrix.

5.4 **Grouping Efficiency (GE)**

Grouping Efficiency can be defined in Equation (3) [21]-[22]:

\[
GE = \rho \frac{N1}{\sum_{k=1}^{K} m_k n_k} + (1 - \rho) \left[ 1 - \frac{NE}{MN - \sum_{k=1}^{K} m_k n_k} \right]
\]  

(3)

Where, MN: refers to the (0-1) matrix size; NE: denotes the number of exceptional elements; N1: refers to the number of 1s inside the clusters; k: denotes the number of clusters; m: refers to the number of machines in kth group; n: is the number of parts in kth group; ρ: is the weight factor ranging between 0 and 1, usually 0.5 is used widely.

5.5 **Machine Utilization (MU)**

Machine utilization refers to the percentage of utilizing the machines inside the cells in the production. [21, 22] proposed Equation 4 to compute MU as follows:

\[
MU = \frac{N1}{\sum_{k=1}^{K} m_k n_k} \times 100
\]  

(4)

Where, N1: denotes the whole number of one’s inside clusters; K: is the number of groups; m: is the number of machines in the kth group; n: is the number of products in the kth group.

6. **Results and discussion**

To validate the performance of the proposed methods, a set of five benchmark problems was chosen from the open literature and has been solved. Firstly, all the test problems were solved by the proposed procedures of the selected methods in sections 4.1, 4.2 and 4.3. Secondly, a number of
machine cells, C, was identified. Furthermore, the performance parameters in sections 5.1 and 5.2 were applied to all the tested problems. The sizes and references of the selected benchmark datasets, in addition to the results of a number of machine cells C, a number of exceptional elements EE, and a number of voids V, are shown in Table 2.

Table 2. The results of C, V, and EE of the three proposed methods (H-H, S-S, S-H) for the selected datasets.

| Dataset | References | Performance Measures |
|---------|------------|----------------------|
|         |            | Hamming-Hamming      | SOM-SOM               | SOM-Hamming               |
|         | C | EE | V | C | EE | V | C | EE | V |
| 5*7     | 2 | 2  | 3 | 2 | 2  | 3 | 2 | 2  | 3 |
| 7*11    | 3 | 2  | 7 | 3 | 4  | 10| 3 | 3  | 8 |
| 10*10   | 3 | 0  | 12| 3 | 0  | 14| 3 | 1  | 12|
| 10*15   | 3 | 0  | 5 | 3 | 0  | 5 | 3 | 0  | 5 |
| 8*20    | 3 | 11 | 2 | 3 | 10 | 1 | 3 | 10 | 1 |

Table 2 shows that the number of machine cells (C) is similar in all the three proposed methods. For the number of EE where the lesser number refers to the better solution as it reduces the intercellular movement cost, the hamming distance shows better outcomes than the other two methods. In the term of the voids (V), the lower number of voids refers to the better utilization of the machines, as it decreases the idle times of machines. The recorded results of (H-H) and (S-H) are almost similar and better than (S-S) with slight differences.

The performance parameters in sections 5.3, 5.4, and 5.5, namely PE, MU, and GE, were also applied to the five selected test problems after they had been solved by the proposed methods. The results of this step are shown in Table 3 and Figure 6.
Table 3. The results of the three proposed methods in the terms of PE, MU, and GE for the selected datasets.

| Dataset | References | Performance Measures |
|---------|------------|----------------------|
|         |            | Hamming-Hamming      | SOM-SOM     | SOM-Hamming |
|         |            | PE  | MU  | GE  | PE  | MU  | GE  | PE  | MU  | GE  |
| 5*7     | [23]       | 0.1250 | 0.8235 | 0.8500 | 0.1250 | 0.8235 | 0.8561 | 0.1250 | 0.7777 | 0.8300 |
| 7*11    | [24]       | 0.0952 | 0.7307 | 0.8457 | 0.1904 | 0.6296 | 0.7748 | 0.1428 | 0.6923 | 0.8167 |
| 10*10   | [25]       | 0.0000 | 0.6571 | 0.8285 | 0.0000 | 0.6216 | 0.8108 | 0.0434 | 0.6470 | 0.8159 |
| 10*15   | [20]       | 0.0000 | 0.9000 | 0.9500 | 0.0000 | 0.9000 | 0.9500 | 0.0000 | 0.9200 | 0.9600 |
| 8*20    | [21]       | 0.1800 | 0.9615 | 0.9300 | 0.1639 | 0.9807 | 0.8941 | 0.1639 | 0.9807 | 0.9440 |

Table 3 and Figure 6 refer to the following outcomes: in terms of the PE, [26] reported that the lower value of PE refers to the better clustering results, the PE results were approximately equal for all the three proposed methods. Still, the (H-H) recorded the highest values of MU followed by (S-H), then (S-S), here the higher value of MU refers to superior clustering results [26].

On the other hand, the H-H based method has an almost more beneficial solution of GE, followed by the S-H-based method and finally, the S-S-based method, where the higher number denotes the better result. In conclusion, the results of the three methods converge, with a slight difference in the H-H-based method.

7. Conclusions:

The main intention of this paper was to assess two methods used for creating MCs and PFs in CMS. These methods are Hamming distance- and SO-based methods. The computational results of the proposed methods and comparison for five benchmark problems were introduced. The results that were obtained from the two selected methods after evaluation process are:

1. The H-H showed better outcomes than the other two methods in terms of the percent of exceptional elements (PE).
2. In terms of voids (V), H-H recorded the best solutions, followed by S-H, then S-S.
3. Likewise, H-H recorded better solutions in terms of machine utilization (MU), followed by S-H, then finally S-S.
4. Furthermore, H-H recorded better solutions in terms of grouping efficiency (GE), followed by S-H, then finally S-S.

In conclusion, H-H showed better solutions than S-H and S-S for the majority of the performance parameters. For future work, it is suggested that:

1. Assessment of other methods, such as array-based methods, similarity coefficient-based methods, meta heuristic-based methods, etc. is conducted.
2. There should be evaluation of stages other than the design stage, for example cell layout, cell schedule.
3. Other performance parameters, such as Intercellular moves (IC), grouping efficacy (GC), group technology efficiency (GTE), etc. could be used.
4. It may be helpful to use another types of matrix rather than 0–1-based matrix within the production factors such as machine capacity, production volume, production time, etc.

![Table showing PE results for different datasets and methods](image-url)

(a): The PE results of the three proposed methods

![Table showing MU results for different datasets and methods](image-url)

(b): The MU results of the three proposed methods

![Table showing GE results for different datasets and methods](image-url)

(c): The MU results of the three proposed methods

![Bar chart showing PE results for different datasets and methods](chart-url)

(d): The GE results of the three proposed methods
Figure 6. The results of the three proposed methods (H-H, S-S, S-H) for the selected datasets.

References

[1] Su M-C, Liu T-K and Chang H-T 2002 Improving the Self-Organizing Feature Map Algorithm Using an Efficient Initialization Scheme, *Tamkang Journal of Science and Engineering* 5 (1) 35-48.

[2] Venkumar p and Haq A N 2006 Complete and fractional cell formation using Kohonen self-organizing map networks in a cellular manufacturing system, *International Journal of Production Research* 44 (20) 4257-4271.

[3] Chattopadhyay M, Dan P K and Majumdar S 2010 Application of visual clustering properties of self organizing map in machine-part cell formation *Applied Soft Computing* 1-33.

[4] Chattopadhyay M, Dan P K and Majumdar S 2011 Principal component analysis and self organizing map for visual clustering of machine-part cell formation in cellular manufacturing system, *Systems Research Forum (SRF)* 5 (1) 25-51.

[5] Chattopadhyay M, Chattopadhyay S and Dan P K 2011 Machine-Part cell formation through visual decipherable clustering of Self Organizing Map, *International Journal of Advanced Manufacturing Technology* 52 (9–12) 1019–1030.

[6] Hamza, S A and Adesta EYT 2013 Integration of the Basic Decisions of the Design of Cellular Manufacturing System, *International Journal of Management Theory and Application* [IREMAN] 1 (6) 354-360.

[7] Chattopadhyay M, Dan P K and Majumdar S 2014 Comparison of visualization of optimal clustering using self-organizing map and growing hierarchical self-organizing map in cellular manufacturing system, *Applied Soft Computing* 22 528-543.

[8] Potočnik P, Berlec T, Sluga A and Govekar E 2014 Hybrid Self-Organization Based Facility Layout Planning *Journal of Mechanical Engineering* 60 (12) 789-796.

[9] Pradhan T and Mishra S R 2015 Implementation of Machine Part Cell Formation Algorithm in Cellular Manufacturing Technology Using Neural Networks, *International Journal of Hybrid Information Technology* 8 (2) 173-178.

[10] Wu L and Suzuki S 2015 Cell formation design with improved similarity coefficient method and decomposed mathematical model Int. J. Adv. Manuf. Technol. 79 1335-1352.

[11] Kiriş S B and Tüysüz F 2017 Performance comparison of different clustering methods for manufacturing cell formation, *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi* 21 (5) 1031-1044.

[12] Hamza S A 2018 Integration of the Assessment and Design of Cellular Manufacturing System *Journal of University of Babylon Engineering Sciences* 26 (4) 316-330.

[13] Tran T-T-Q, Phan T-C, Laurent A, D’Orazio L 2018 Improving Hamming distance-based fuzzy join in MapReduce using Bloom Filters *FUZZ-IEEE: International Conference on Fuzzy Systems* (Rio de Janeiro, Brazil. hal-01857386).
[14] Hamza S A and Jehad A 2019 Heuristic method for solving cell formation problem in cellular manufacturing system based on hamming distance The Iraqi Journal for Mechanical and Materials Engineering 19 (1) 75-90.

[15] Haykin S 2009 Neural networks and learning machines Third edition, Copyright 2009 by Pearson Education, Inc. (Upper Saddle River, New Jersey 07458).

[16] Kohonen T 2014 MATLAB Implementations and Applications of the Self-Organizing Map (Unigrafia Oy, Helsinki, Finland).

[17] Tang M, Yu Y, Aref WG, Malluhi QM and Ouzzani M 2015 Efficient Processing of Hamming-Distance-Based Similarity-Search Queries Over MapReduce. Proc. 18th. Int. Conf. on Extending Database Technology (EDBT) March 23-27, (Brussels, Belgium).

[18] Wang J, Shen H T, Song J and Ji J 2014 Hashing for Similarity Search: A Survey arXiv:1408.2927 V1 [cs.DS].

[19] King J R 1980 Machine-component grouping in production flow analysis: An approach using rank order clustering algorithm International Journal of Production Research 18 213-232.

[20] Chan H M and Milner D A 1982 Direct clustering algorithm for group formation in cellular manufacture Journal of Manufacturing Systems 1 64-76.

[21] Chandrasekharan M P and Rajagopalan R 1986a An ideal seed non-hierarchical clustering algorithm for cellular manufacturing International Journal of Production Research 24 451-464.

[22] Chandrasekharan M P and Rajagopalan R 1986b MODROC: An extension of rank order clustering for group technology International Journal of Production Research 24 1221-1233.

[23] Waghodekar P H and Sahu S 1984 Machine-component cell formation in group technology: MACE International Journal of Production Research 22 (6) 937–948.

[24] Boctor F F 1991 A linear formulation of the machine-part cell formation problem. International Journal of Production Research 29 343-356.

[25] Mosier C T and Taube L 1985 Weighted similarity measure heuristics for the group technology machine clustering problem Omega 13 (6) 577–583.

[26] Chu C-H and Tsai M T 1990 A comparison of three array-based clustering techniques for manufacturing cell formation International Journal of Production Research 28 (8) 1417-1433.