

# Machine Learning Approach to Community Detection in a High-Entropy Alloy Interaction Network

Raheleh Ghouchan Nezhad Noor Nia, Mehrdad Jalali,\* Matthias Mail, Yulia Ivanisenko, and Christian Kübel



Cite This: *ACS Omega* 2022, 7, 12978–12992



Read Online

ACCESS |

Metrics & More

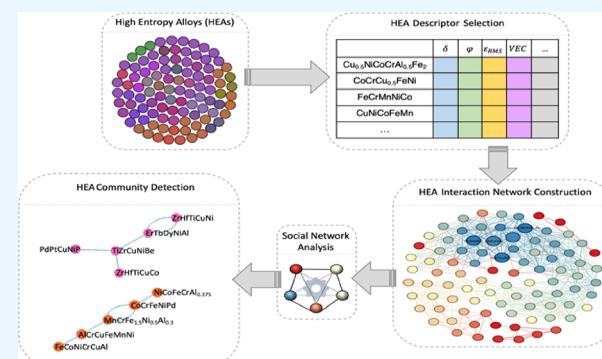
Article Recommendations

**ABSTRACT:** There is a growing trend toward the use of interaction network methods and algorithms, including community-based detection methods, in various fields of science. The approach is already used in many applications, for example, in social sciences and health informatics to analyze behavioral patterns during the COVID-19 pandemic, protein–protein networks in biological sciences, agricultural science, economy, and so forth. This paper attempts to build interaction networks based on high-entropy alloy (HEA) descriptors in order to discover HEA communities with similar functionality. In addition, these communities could be leveraged to discover new alloys not yet included in the data set without any experimental laboratory effort. This research has been carried out using two community detection algorithms, the Louvain algorithm and the enhanced particle swarm optimization (PSO) algorithm. The data set, which is used in this paper, includes 90 HEAs and 6 descriptors. The results reveal 13 alloy communities, and the accuracy of the results is validated by the modularity. The experimental results show that the method with the PSO-based community detection algorithm can achieve alloy communities with an average accuracy improvement of 0.26 compared to the Louvain algorithm. Furthermore, some characteristics of HEAs, for example, their phase composition, could be predicted by the extracted communities. Also, the HEA phase composition has been predicted by the proposed method and achieved about 93% precision.

## 1. INTRODUCTION

Since the ancient eras, human civilization has attempted to discover new and unknown materials, for example, metals and alloys that can all play a key role in the overall quality of human life. Since the Bronze Age, alloys have been produced based on a “basic element” pattern containing one principal element. Various elements are added to the basic element to improve selected properties.<sup>1</sup> Over the past decades, a new approach has been introduced to design alloys, which involves mixing typically five elements or more in equimolar amounts to produce balanced alloys called high-entropy alloys (HEAs).<sup>1</sup> These were initially introduced and developed by Cantor et al.<sup>2</sup> and Yeh et al.<sup>3</sup> The entropy of mixing for these complex alloys is high. The atoms used to create the alloys have a similar size.<sup>6</sup>

HEAs have been widely investigated due to their attractive properties, for example, thermal and electrical conductivity,<sup>4</sup> high corrosion resistance,<sup>5</sup> and high strength in combination with high ductility. A parametric method is commonly used to understand and predict the phase stability, often used in pairs presented by a two-dimensional diagram.<sup>17</sup> Although there are many parametric and statistical methods in the field of



materials science, machine learning (ML) is considered one of the most effective methods in materials science.<sup>7</sup> ML algorithms are capable of learning models to explore communities and provide results effectively.

The purpose of the present study is to introduce a new model for the HEA interaction network, which is made based on HEA descriptors. This model measures similarities among HEA descriptors by creating a network of interactions based on similarities. Communities are extracted from the interaction network so that each community comprises similar HEA compounds. These compositions are available for interpretation. The outcomes of this paper are communities that can help anticipate HEA phases and detect HEA functionalities through the ability to better analyze them. With that, it might

Received: January 16, 2022

Accepted: March 7, 2022

Published: April 4, 2022



be possible to suggest more efficient alloys for selected applications.

## 2. REVIEW OF LITERATURE

The HEAs containing at least five elements with equal or similar atomic percentages have high strength.<sup>32</sup> These alloys are different from conventional alloys due to four main effects, which include the (1) high-entropy effect, (2) sluggish diffusion effect, (3) lattice distortion effect, and (4) cocktail effect.<sup>32</sup> These effects contribute to the ultimate strength or hardness of HEAs.<sup>32</sup>

In a study carried out by Ye et al.,<sup>1</sup> the phase formation of HEAs and their new properties are discussed, such as strength, mechanical performance at high and cryogenic temperatures, ductility, hardness, magnetism, and electrical conductivity.

By using ML in HEAs, the design of alloys can be facilitated and used to discover new compounds with desirable properties. Dai et al.<sup>8</sup> introduced a method that creates a low dimensional descriptor for predicting the phase content of HEAs based on their composition. The process behind their method has several stages: first a coefficient analysis is used to select closely related highly relevant descriptors. It then increases dimensions, which is based on the main structure of primary functions and finally, it is important to select descriptors to explain the material.<sup>8</sup> The main focus is on predicting the alloy phase composition. Using a Pearson correlation, the highly correlated descriptors are removed. They created a new nonlinear descriptor, which analyzes the relationship between descriptors to eliminate additional descriptors.<sup>8</sup> Based on the study performed, the authors proposed a framework to collect HEAs data for interpretation.<sup>27</sup> The features are removed which are causing weak phase predictions.<sup>27</sup> The various ML classifiers are used to predict the HEAs phases the HEA interaction network is not created to extract similar compounds as a community.<sup>27</sup> In a study performed by Kaufmann and Vecchio,<sup>21</sup> a ML method is applied to HEAs to predict solid solution forming ability. Thermodynamic and chemical features are used to do predictions by a random forest model.<sup>21</sup> The HEAs descriptors are not used to compute a hybrid similarity to construct a HEA interaction network.<sup>21</sup> identify similar Based on another study, the authors presented a structure by using a genetic algorithm to choose an effective ML model and features for HEAs phase forecast.<sup>22</sup> The content and structural similarity are not considered by the proposed technique to make a HEAs interaction network.<sup>22</sup> Similar HEAs compositions are not detected as a cluster.<sup>22</sup> To the best of our knowledge, a full interaction network of HEAs has not yet been created to analyze it by social network tools. One of the social network analysis tools is community detection. Communities are a group of compounds that can be used to improve functionality, new compound, and new descriptor discovery. Most studies are about phase detection of alloys by ML methods, which predict the phase composition of a compound, HEA communities can be extracted to predict phases more accurately. In addition, each community consists of HEAs with a similar descriptor and behavior.<sup>26</sup>

In a study, carried out by M'Barek et al.,<sup>26</sup> biological interaction networks have explored such as genes or proteins. The communities extracted from these biological networks are a set of proteins or genes that collaborate on a similar cellular functionality. By using a genetic algorithm, they presented a specific fitness function based on the amount of similarity and

interaction between genes.<sup>26</sup> They have used the semantic similarity in a KEGG data set, which they have taken as score for the structure of communities. It is calculated according to the semantic similarity method based on a genetic ontology.<sup>26</sup> Based on another study, a modified Deepwalk algorithm is presented by the authors, which predicted a link in the protein-protein interaction network.<sup>31</sup> The feature dimensions are reduced to integrate the network structure and features for link prediction,<sup>31</sup> an HEA network by using their descriptors that is used to detect community for similar nodes.<sup>31</sup>

In another study, the authors proposed a new tool named MOFSocialNet based on creating social networks using a metal–organic framework (MOF) database. MOFSocialNet is able to guide MOF researchers through the vast chemical space of existing and hypothetical MOFs. For a demonstration, they used social network analysis to identify the most representative MOFs in this research data set and to detect MOF communities.<sup>33</sup>

In another study carried out by Ahajjam et al.,<sup>25</sup> a scalable and deterministic approach is proposed to identify communities using leader nodes called the community leader recognition approach. Their approach has two main steps: the first step is to retrieve the leaders and the second step is to identify the community using the similarities between the nodes. Two important issues in their work are community recognition and leader detection in complex networks.<sup>25</sup> The network leader nodes are responsible for disseminating the influence and then, using the similarities between the nodes, the communities around the leader are formed. In social networks, the central nodes are responsible for spreading the intrusion. The advantage of this method is that there is no need for prior knowledge of the number of leaders and communities. They start by finding a leader to identify the most effective nodes and then extract the communities. For each leader, a community is obtained by calculating the similarity between the nodes.<sup>25</sup> They distinguish communities based on the similarities of the nodes with the leader, who are all in the leader's neighborhood. They used real social network data sets and used the Jaccard, Salton, Human Development Index (HDI) and Human Poverty Index (HPI) to calculate the similarity for finding out which works best in finding the leader of their method.<sup>25</sup>

In a study carried out by Zhao et al.,<sup>28</sup> a community detection algorithm based on graph compression is introduced, which is effective in large networks. The compressed graph is first obtained by repeatedly merging nodes of degree one or two with their bigger degree neighbors. Then, two indexes, namely, the density and quality of nodes, are defined to evaluate the probability of nodes as the seed of a community. With these two criteria in a compressed social network, the number of communities and the initial members of the related community are determined.<sup>28</sup> They use the real social network data set to evaluate their method. There is no similarity computation between nodes to create a HEA interaction network.

In a study conducted by Rostami et al.,<sup>30</sup> a community detection based on a genetic algorithm is presented for feature selection, which has three steps. It first calculates the similarities of the features, then the features are classified into clusters by community detection algorithms, and finally the features are selected by a genetic algorithm with repair operation based on a new community. A community detection method is used in their approach to divide features into

different groups. Using Pearson correlation similarity, the similarity of the features is examined.<sup>30</sup> Clustering is performed on features, and a threshold is set for determining the number of features in each cluster to reach that number using random repair or score repair. Their proposed method selects the optimal number of features, which is automatically determined based on the overall feature structure and their internal similarities.<sup>30</sup> They achieved accurate results in community detection by the genetic algorithm.<sup>30</sup>

In a study conducted by Ozaki et al.,<sup>20</sup> the pruning method was added to the Louvain algorithm to optimize computational time while maintaining the quality of the community detection and its process. Using this method simplifies the entire process as calculations on the quality of the clusters do not occur at each phase for all nodes, but instead, such calculations are done for nodes that are used in the next phase of the community detection. Ozaki et al.<sup>20</sup> have applied the Louvain algorithm to network that have similarity among nodes calculated based on a Cosine similarity.<sup>20</sup>

The novelty of this paper lies in establishing an interaction network for HEAs that has not been implemented in the field of alloy metallurgy so far. In addition, an interaction network analysis method has been used to analyze the HEAs. This particular method uses ML algorithms for the alloy community detection, along with the Louvain algorithm and the particle swarm optimization algorithm.

### 3. PROPOSED METHOD

In this section, we present our approach for community detection based on interaction networks of HEAs, using the concepts of Louvain and modified particle swarm optimization (PSO) algorithms. The community members are HEAs that are merged to find the best community number by the Louvain method and by the modified PSO, which are considered the node connections. Our method solves the community detection problem by maximizing an objective function called modularity.

Initially, a descriptor for HEAs is selected and a preprocessing is carried out. Communities are extracted by Louvain and by modified PSO algorithms. Our approach is based on the following five steps:

- Data set was preprocessed to perform ML algorithms.
- Descriptor content similarity was calculated.
- Interaction network of HEAs was created.
- Descriptor structural similarity was calculated.
- Communities that maximize the objective function were extracted

The data set used in this paper contains 90 HEA alloys, as listed in Table 1 in ref 1, each of which is characterized by six descriptors (Appendix A).<sup>1</sup>  $\delta$  is the atomic size difference.<sup>9</sup>  $\Delta H_{\text{mix}}$  (in kJ/mol) is the mixing enthalpy, calculated using eq 1:<sup>1</sup>

**Table 1. Comparison of Quality of Community Detection and Extraction Using Modularity Criteria with the Developed Louvain Algorithm and the Particle Swarm Optimization Algorithm**

community detection algorithm	modularity criteria
developed Louvain algorithm	0.7130
PSO algorithm	0.8912

$$\Delta H_{\text{mix}} = \sum_{i=1, i \neq j}^n \Omega_{ij} c_i c_j = \sum_{i=1, i \neq j}^n 4 \Delta H_{ij}^{\text{mix}} c_i c_j \quad (1)$$

$S_c$  (in  $k_B/\text{atom}$ , where  $k_B$  is Boltzmann's constant) is the configurational entropy of mixing for an ideal solid solution.<sup>1</sup>  $\varphi$  is a single dimensionless thermodynamic parameter for designing HEAs.<sup>6</sup>  $\epsilon_{\text{RMS}}$  is the root-mean-square residual strain, usually measured through the energy storage density of the elastic pressure.<sup>10</sup> VEC is considered an important parameter in the selection of the valence electron concentration of the alloys due to the lack of robust atomic size difference.<sup>1</sup> It should be noted that the data class label can also be called phase, which is not considered in these calculations, and the most important challenge in the current article is to find the relationship among similar alloy types. The six descriptors and a portion of the data set are listed in Table A1 in Appendix A, the first column of which contains the number for each chemical composition of the HEA alloy used in the results in Section 5. The second column in Table A1 gives the HEA chemical composition, and the other six columns show the values of the six descriptors for each composition.

Algorithm 1 shows the pseudocode of the proposed method that detects communities using Louvain and PSO algorithms:

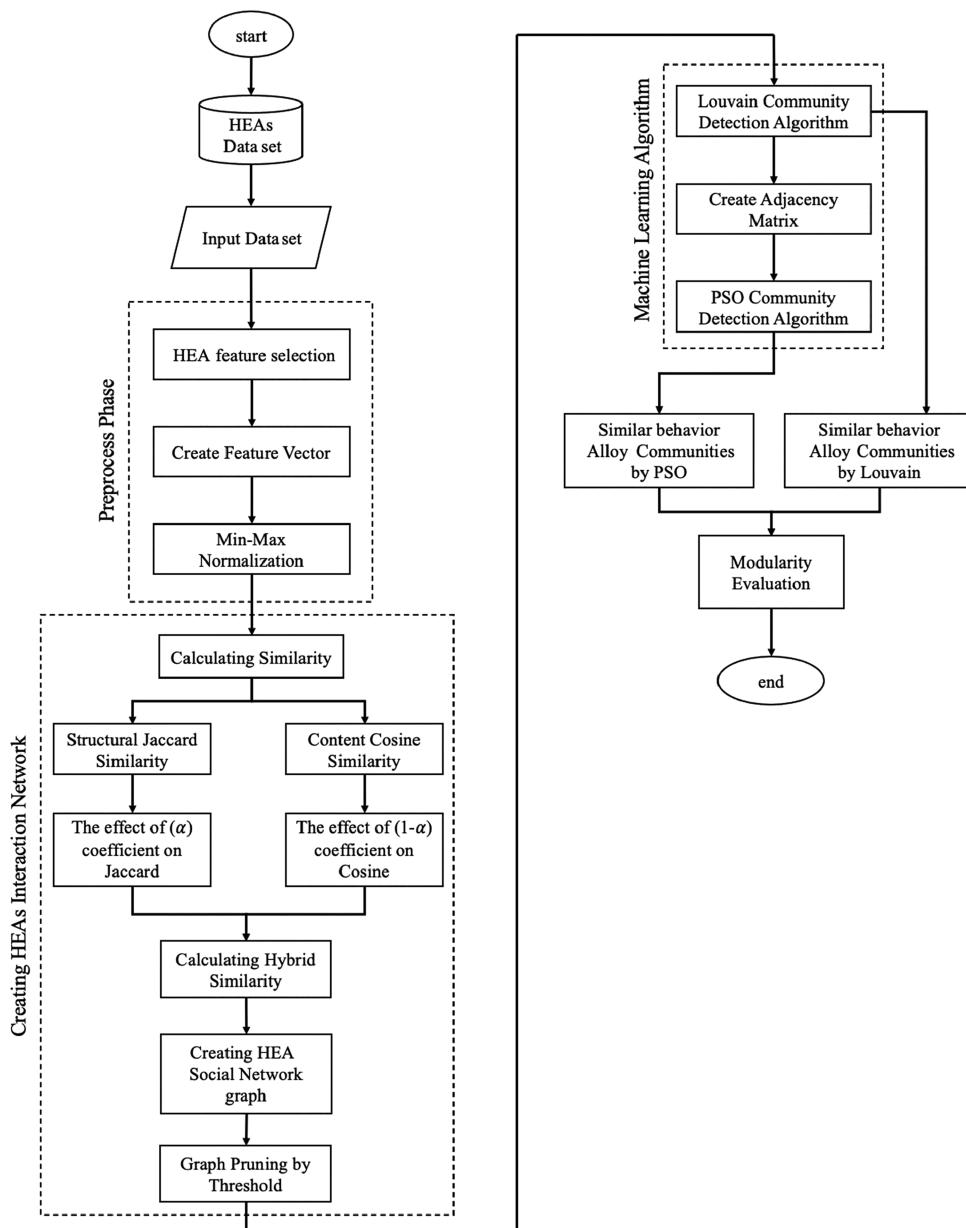
```

Input: undirected, weighted HEAs Interaction network graph G=(V,E)
Output: Set C=(C1, C2, ..., Cn)
1: Preprocess: Select 6 descriptors of HEAs
2: Create descriptor vector
3: Normalize descriptors value by Min-Max method using Formula (2) to transfer to range [0,1]
4: HEAs Interaction network creating: Find the similarities of each node i and j using Structural Jaccard Similarity Formula (4) and content Cosine Similarity Formula (5)
5: Calculate Hybrid Similarity by considering Alpha coefficient that Hybrid Similarity= ( $\alpha * (\text{Jaccard Matrix}) + ((1 - \alpha) * (\text{Cosine Matrix}))$ )
6: Connect each vertex V to each other to create HEAs Interaction network
7: Pruning Interaction network graph by Threshold that
8: if (Vertex similarity >= Threshold)
    Vertex Similarity=1
9: else
    Vertex Similarity=0
10:
11:
12: Machine Learning Algorithm: Louvain Algorithm (Graph G)
13: Calculate adjacency(G,'weighted')
14: C the index of community of each node of G'
15: Initialize each node with its own community
16: C=Move nodes (G')
17: G'=Aggregate(G',C)
18: C=put each node of G'
19: return G' and communities
20: Evaluate Community quality by Modularity measurement using Formula (5)
21: Use full Adjacency that obtained from step 13 to PSO Algorithm
22: Input the HEAs array to create Initial population in PSO
23: Calculate PSO fitness for Community Quality using Formula (9)
24: Determine the node location and velocity using Formulas (10), (11) and (12)
25: return Modularity measurement and communities

```

The flowchart of the proposed method is shown in Figure 1. There are three stages in the proposed method. The first stage is preparing the data, which consists of three steps including HEA feature selection, feature vector creation, and normalization. The second stage is creating an HEA interaction network by using similarity and pruning graph methods. The third step is to apply a ML algorithm that extracts communities from the network. Finally, the modularity is measured, which shows the quality of the communities.

**3.1. Data Normalization.** Normalization is used when the provided data values are not in the same range and have different intervals to prevent properties and descriptors that contain large values to dominate the overall performance of the system. Additionally, the normalization can potentially minimize the impact of out-of-range scales and maintains all inputs in a single interval. In the present article, min–max normalization was used for property values to normalize the property values to the interval [0, 1] using eq 2:<sup>13</sup>



**Figure 1.** Flowchart of the proposed method with details. The process is done in three phases including the preprocess, creation of the HEA interaction network, and ML algorithms.

$$\nu' = \frac{\nu - \min_A}{\max_A - \min_A} \quad (2)$$

where  $\min_A$  and  $\max_A$  indicate the current minimum and maximum values of the properties found in A. The original values and the normalized values of the properties are presented as  $\nu$  and  $\nu'$ , respectively. As can be seen in eq 2 used above, the maximum and minimum values are 1 and 0.<sup>13</sup>

**3.2. Content Cosine Similarity Criteria.** Content cosine similarity is measured based on the internal angle between two vectors and determines whether the selected vectors are considered codirectional.<sup>11</sup> As shown in the data set in Appendix A, each property of a single composition can be analyzed and compared to another compound.<sup>1</sup> Equation 3 shows the content cosine similarity as follows<sup>11</sup>

$$\text{content cosine similarity } (x, y) = \frac{\sum_i (x_i y_i)}{\sqrt{\sum_i x_i^2 \sum_i y_i^2}} \quad (3)$$

where  $x_i$  represents the  $i$ th property of the first compound and  $y_i$  is the  $i$ th property of the second compound.

**3.3. Structural Jaccard Similarity Criteria.** The Jaccard index is mainly used for a comparison of the structural similarity of a data set.<sup>12</sup> The value of the Jaccard similarity coefficient between two data sets is usually obtained by dividing the number of common properties of the two available sets by the total properties of the two sets.<sup>12</sup> Because the input of the interaction network graphical structure is required for the calculation of the Jaccard criterion, the matrix obtained by content cosine similarity must be examined first with different thresholds to find the appropriate value and create the desired graphical representation of the network, so that structural similarities can be measured using the established graph. The

threshold to obtain a graph for the content cosine similarity analysis was set at 0.98 in the current study.

The description of the structural Jaccard similarity criteria is shown in eq 4:<sup>12</sup>

$$\text{structural Jaccard similarity } (v_i, v_j) = \frac{|N_i \cap N_j|}{|N_i \cup N_j|} \quad (4)$$

where  $v_i$  and  $v_j$  are the two nodes representing the compounds of HEAs,  $|N_i \cap N_j|$  denotes the common properties of the two compounds  $v_i$  and  $v_j$ , and  $|N_i \cup N_j|$  are all the properties of  $v_i$  and  $v_j$ . It is also important to note that this particular criterion can be applied to all common pairs of attributes.

**3.4.  $\alpha$  Coefficient.** The calculation of the parameters for content and structural similarity results in two matrices with similarity values. To detect communities, a hybrid similarity matrix is needed as input that contains similar properties. The  $\alpha$  coefficient determines the effect of each of the similarities. The  $\alpha$  coefficient also determines the possible effect of each of these common similarities as well as the effect of structural Jaccard similarity. The  $\alpha$  coefficient also determines the proper effect of each of content and structural similarities. The output of this phase is a hybrid similarity matrix as required for the community detection algorithm.

#### 4. COMMUNITY DETECTION

Each community in the interaction network shows the alloys that have dependencies between each other to perform the same functionality in an equal community.

**4.1. Theoretical Notation Definitions.** A complex network can be mapped to the graph  $G(V,E)$ , where  $V$  is the node set and  $E$  is the edge set. A network  $C(v,e)$  is said to be subnetwork, if  $v$  is the subset of  $V$  and  $e$  is the subset of  $E$ .

Let  $A$  be the adjacency matrix; two nodes are adjacent if they have an edge between them. If there exists a link between vertex  $i$  and vertex  $j$ , then  $A_{ij} = 1$ ; otherwise  $A_{ij} = 0$ . A weighted network has weight  $w$  joined to the edges, where  $w$  is a real number.

Communities in networks are the groups of nodes, which are more profoundly connected to each other than to the rest of the nodes within the network. Community detection is the key characteristic, which may well be utilized to extract valuable information from networks.<sup>29</sup>

**4.2. Louvain Algorithm.** The content of science studies can usually be represented as complex networks, in which the topology of interconnected vertices is obtained from either an organized or random compound.<sup>14</sup> The Louvain algorithm is a metaheuristic method that is introduced to identify and detect communities and groups within the provided graph. In addition, each extracted group represents a community, and this type of algorithm is considered to be an ascending clustering method.<sup>15</sup> Furthermore, a parameter called modularity is used to determine the quality of the obtained communities in this algorithm, and the maximization of this particular parameter is considered to be of great importance. This specific parameter selects the type of communities that are integrated with the target vertex and creates highly modular communities.<sup>15</sup> Despite difficulties in the calculation of modularity in large graphs, the Louvain algorithm can overcome this issue by speeding up the processing of large graphs.<sup>15</sup> This unique property of the algorithm led to its popularity.<sup>16</sup> It is also essential to add that the Louvain algorithm is considered the fastest and most effective algorithm

for community detection that tends to operate tirelessly to achieve maximum modularity. The implemented algorithm is divided into two phases that are alternately repeated. Imagine that the procedure begins through a weighted interaction network with  $N$  nodes. It first places each node in a separate community, which has just as many nodes as the current network, and then examines the possible neighbors for each node and evaluates the precise rate of modularity, which is accomplished through the removal of the nodes from its related community and transferring them to its neighboring community. Finally, the targeted nodes are placed within the community with the highest possible modularity rate (positive rate); otherwise, it will remain in the current community. Afterward, this process is repeated alternately for all the interaction network nodes of HEAs until no new enhancements are achieved and the first stage of the process is essentially completed.<sup>17</sup> Although this process is repeated several times for each node, the first stage is completed when a local maximum modularity is reached and the rate of modularity remains stable. Examining the order of the nodes in the output of the algorithm may affect the computational load that requires further study.

The overall performance quality of the Louvain algorithm for the community detection can be obtained using the modularity rate  $\Delta Q$ , which is calculated through transferring isolated node  $i$  to the  $C$  community via eq 5:<sup>17</sup>

$$\Delta Q = \left[ \frac{\Sigma_{in} + k_{i,in}}{2m} - \left( \frac{\Sigma_{tot} + k_i}{2m} \right)^2 \right] - \left[ \frac{\Sigma_{in}}{2m} - \left( \frac{\Sigma_{tot}}{2m} \right)^2 - \left( \frac{k_i}{2m} \right)^2 \right] \quad (5)$$

where  $\Sigma_{in}$  is the sum of the links found within the community  $C$ ,  $\Sigma_{tot}$  is the total weight of the links connected to the nodes within the community  $C$ ,  $k_i$  is the total weight of links associated with node  $i$ ,  $k_{i,in}$  is the sum of the weights of the links connected from node  $i$  within the community  $C$ , and  $m$  is the total weight of all the network links. When node  $i$  is transferred from its related community, a similar term is often used to evaluate the modularity changes and adjustments, and the modularity changes can be measured through the removal of node  $i$  from its related community and its replacement in the neighboring community.<sup>17</sup>

The second stage of this particular algorithm involves establishing a new network of nodes that have previously found their community during the first phase. The weight of the links among new nodes is reached through the total weight of the links between nodes in two respective communities, and the links between the nodes in the same community can potentially lead to an inner circle of the community in the newly established networks. After the second phase is completed, the initial phase of the algorithm can then be re-applied in the previously created weighted network to evaluate the obtained results more accurately, and the combination of these two phases is termed as a pass. As a result, the overall number of meta-communities decreases with each iteration, and its highest computational load occurs in the initial phase. In fact, these phases are to be continued until the maximum modularity is reached and no further changes occur. This particular algorithm can represent highly complex networks and often operates hierarchically so that the final obtained

communities are created and established through an iterative process of integration.

Moreover, the height of the hierarchy is determined through the number of iterations, which is usually discovered to be small; take note that this algorithm can possibly have various advantages, such as the visibility of courses that can easily be conducted as well as reaching the targeted outcome without a need for individual attendance or any type of monitoring. Next, it should also be added that the algorithm operates quite as fast and can calculate the modularity rate simply based on [Formula 5](#), which after several repeated courses tends to reduce the number of obtained communities through the integration method. The maximum conduction period of this particular method is related to the initial iteration in the first phase.<sup>17</sup> Also, the qualitative limitations and boundaries of modularity have been eliminated, due to the multilevel nature of the algorithm. Finally, it is also worth mentioning that the isolated nodes are transferred from one community to another, in the first phase of the algorithm.<sup>17</sup> The probability of merging two separate communities through transfer of nodes one-by-one is considered extremely low. However, take note that these communities can very well be merged later, after the consolidation of the nodes is complete.<sup>19</sup>

**4.3. Community Detection Based on the Particle Swarm Optimization Algorithm.** Kennedy and Eberhart initially introduced the PSO algorithm in 1995, which was inspired by the characteristics and behavior of birds.<sup>24</sup> The PSO algorithm is considered to be one of the most important and useful swarm intelligence algorithms, which frequently offers better overall solutions compared to other available algorithms. The mobility found within the particles, which is an array ( $90 \times 1$ ) of nodes, is potentially the best possible way to update each particle for community detection.<sup>19</sup> Optimization of the algorithm may lead to rapid convergence as well as a reduction in the rate of references to the proportionality function, which is directly related to the modularity criterion of community quality.<sup>19</sup> For example, suppose there is a major optimization issue found in the dimensional space of  $d$ , where  $X_i = (X_{i1}, X_{i2}, \dots, X_{id})$  and  $V_i = (V_{i1}, V_{i2}, \dots, V_{id})$  are the position and velocity vectors, respectively. Let  $pbest_i$  be the best possible solution for particle  $i$  ( $i = 1, 2, \dots, P_{size}$ ) and  $gbest$  be the best possible solution among any type of particle. Furthermore, collaborative and ML of particles are also conducted in each update of  $pbest_i$  and  $gbest$ . Besides, with each iteration of the PSO algorithm, the current velocity and position of the particles can also be updated using [eqs 6–8](#) as follows:<sup>23</sup>

$$\begin{aligned} V_i(t+1) &= wV_i(t) + c_1 + rand_1[pbest_i - X_i(t)] + c_2 \\ &\quad + rand_2[gbest - X_i(t)] \end{aligned} \quad (6)$$

$$X_{ij}(t+1) = \begin{cases} 1, & \text{if } \rho < sig(V_{ij}(t+1)) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$sig(V_{ij}(t+1)) = 1/[1 + exp(V_{ij}(t+1))] \quad (8)$$

in which the parameter  $t$  represents the iteration of the conducted algorithm,  $w$  is essentially the inertia coefficient,  $c_1$  and  $c_2$  are the learning rates,  $rand_1$  and  $rand_2$  are random numbers that are uniformly generated in the interval [0, 1], and  $\rho$  generally functions as a predefined threshold.<sup>23</sup>

**4.3.1. Optimal PSO Algorithm and Group Learning.** Given that independent communities are obtained by sorting the set of HEA compounds in the interaction network of  $L(G)$  materials, they are optimal communities and smaller than  $G$ . To identify independent communities in a network, there is a need to discover independent communities in the corresponding line graph. The developed PSO algorithm is used along with group learning techniques resulting in LEPSO, which can be used to optimize the results obtained by linear graph segmentation.<sup>23</sup>

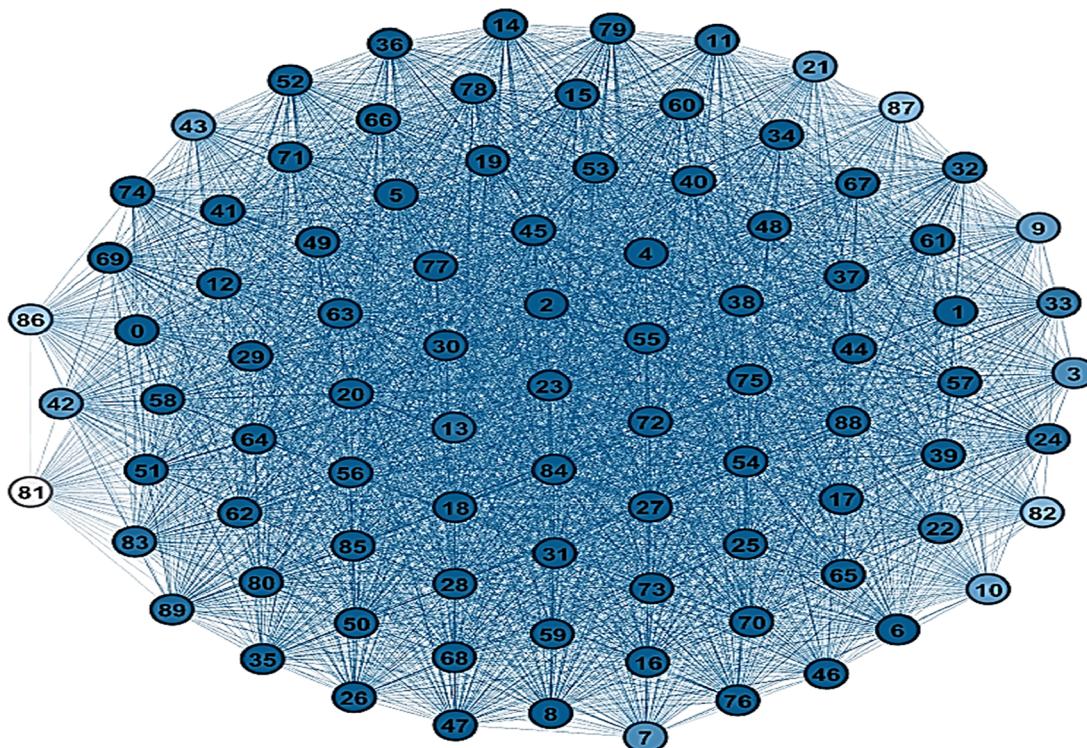
**4.3.2. Presentation of Community Detection Using Optimal PSO.** The linear graph for the chemical composition of alloys is represented as  $L(G) = \langle N, E \rangle$ , where  $N = (n_1, n_2, \dots, n_k)$ , in which a part of  $L(G)$  can be presented as  $X_i = (X_{i1}, X_{i2}, \dots, X_{id})$  and  $k = |N|$ . In the case where the initial value was assumed to be  $X_{ij} = m$ , then the results may indicate that there is a relationship between the two compounds  $e = \langle n_p, n_m \rangle$  and the  $X_i$  particles, specifically when  $n_p$  and  $n_m$  are found within the same communities as  $L(G)$ . In order to determine the initial community as the optimal type, each PSO must first be considered as an array of alloy compounds. In this regard, the matrix proximity of the primary interaction network gains the materials using the connected and linked nodes. Some of the potential drawbacks of this design include random initialization of the particles and frequent updates of the particle locations. Moreover, this issue is often so major that the particle components may potentially display links that have never existed before. To solve such problems, particles are recommended to be presented on a list of regular neighbors.<sup>23</sup>

The foundation of this particular design is essentially based on the use of data distribution of the neighbors for each node as a representative of an alloy composition, which potentially ensures that newly entered particles used in the process of transference or initialization are all allowed. However, the complete removal of unauthorized particles as well as the prevention of the production of local optimal communities, with the use of repetitive binary division and automated community detection methods, are all considered some of the potential advantages of this optimization method in PSO.<sup>23</sup>

**4.3.3. Particle Fitness Function in the Optimal PSO.** The comprehensible definition of community can encourage researchers to introduce new and different types of quality indicators to evaluate the possible benefits of a partition. The main assumption behind modularity is that the edge density of a cluster should be higher than the predicted density of the sub-graph, so the nodes can randomly be linked. In order to complete the discretization process of the provided algorithm, each node and its relationship with the other available nodes has to be individually analyzed and checked. Therefore, the link between the initial compound and the other compositions is obtained first, and the adjacency matrix is established subsequently. Finally, the particle fitness function that can determine the quality of final-phase communities, also known as modularity is shown in [eq 9<sup>23</sup>](#):

$$fit(P_i) = Q(C) = \sum_{c=1}^m \left[ \frac{l_c}{|E|} - \left( \frac{d_c}{2|E|} \right)^2 \right] \quad (9)$$

In this equation,  $fit(P_i)$  is the particle proportionality and fitness value of  $P_i$ ,  $m$  is the number of communities found in the  $C$  partition of the network as  $G = \langle N, E \rangle$ ,  $l_c$  is the number of edges that link the vertices in the community, which is



**Figure 2.** Weighted interaction network of 90 HEAs, where all compounds are connected to each other. The HEA interaction network is fully connected before applying the threshold.

shown as  $c \in C$ ,  $d_c$  is the sum of the nodes within  $C$ , and  $|E|$  is the total number of edges found in  $G$ .

**4.3.4. Update of Particle Speed and Position.** **4.3.4.1. Particle Speed Update.** An optimal particle velocity updating algorithm called GbestGenerator is used to avoid the local optimization method, which applies a voting-based clustering technique to take full advantage of valuable hidden community patterns found in less efficient particles and *gbest* values. In case the proportion of *gbest* does not improve in  $T_{\max}$  consecutive iterations, meaning that if the swarm particles are trapped in, member particle clusters MPS are created through the selection of all available *gbest* particles within the  $T_{\max}$  and its consecutive iteration particles leading to the combination of the right MPS particles to produce new *gbest* particles. Accordingly, each particle can potentially have a minimum and a maximum speed for velocity.<sup>23</sup>

Equation 6 suggests that the inertia coefficient shown as  $w$  is considered extremely significant in the implementation of particle velocity updates. The adjustment strategy of  $w$  can be well expressed using eq 10, which is described as follows<sup>23</sup>

$$w_t = (w_{\max} - w_{\min}) \frac{t_{\max} - t}{t_{\max}} + w_{\min} \quad (10)$$

where  $w_{\max}$  and  $w_{\min}$  are the initial and final inertia coefficients, respectively,  $t_{\max}$  is the representation of the maximum iteration, and  $t$  indicates the current iteration. As can be seen in eq 10, in the initial stage ( $t = 0$ ), the parameters  $w_{\max}$  and  $w_t$  are both considered correspondence of each other. When  $t$  is too close to  $t_{\max}$ ,  $w_t$  gradually decreases toward  $w_{\min}$ . Furthermore, due to the algorithm converges in the early stages, larger coefficient values are needed for the particles to be faster in velocity, while in later stages, much smaller

coefficient values are provided to the particles to gradually enhance their overall stability.

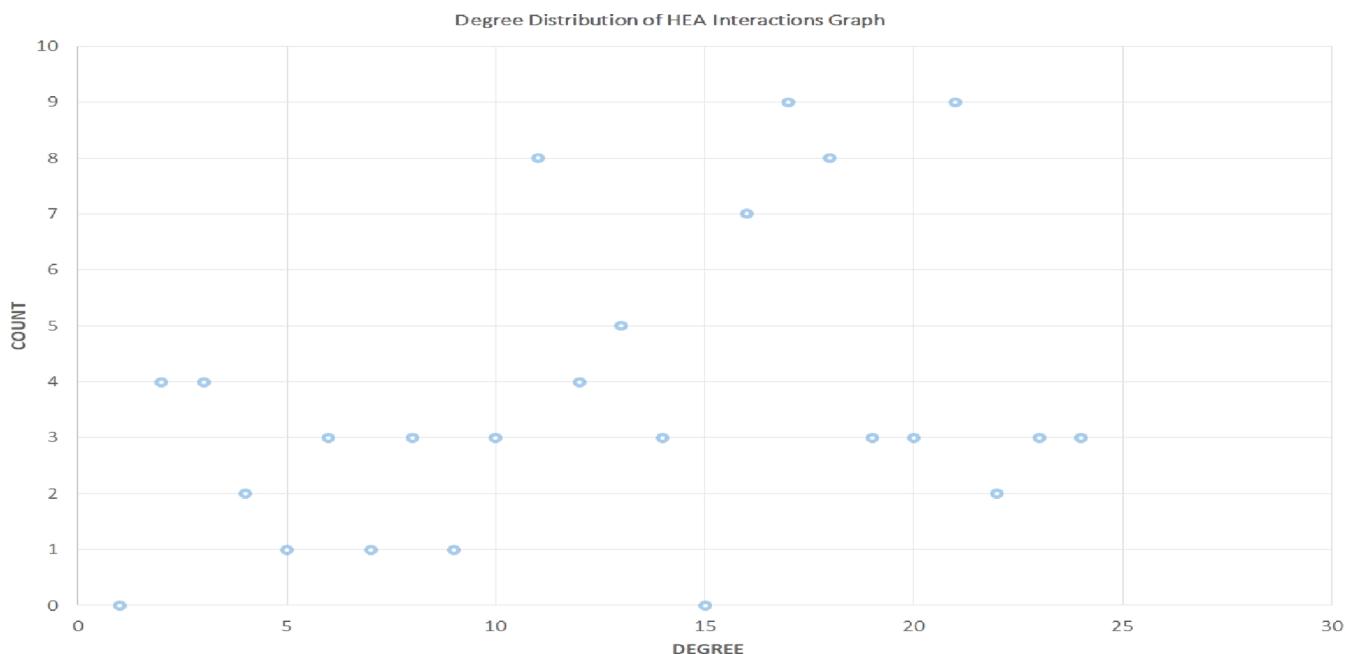
**4.3.4.2. Positional Particle Update.** Based on the previously provided Formula 7, the positional vector components are assigned to either particles 0 or 1, which is not very suitable for the display of particles with respect to the neighbor. Accordingly, the previous position of the particles is related to the previous community and the current new position can be related to the final community. Therefore, the value of  $X_{ij}$  as a part of  $i$  is obtained from an integer within the range from 1 to  $\deg(n_i)$ , meaning that  $X_{ij} \in \{1, 2, \dots, \deg(n_i)\}$  can essentially improve the PSO and the searching abilities of the system. The particle positional updates are shown in great detail using following eqs 11 and 12:<sup>23</sup>

$$X_{ij}(t+1) = \begin{cases} k, & \text{if } [\rho < \text{sig}(V_{ij}(t+1))] \wedge [\deg(n_j) > 1] \\ X_{ij}(t), & \text{otherwise} \end{cases} \quad (11)$$

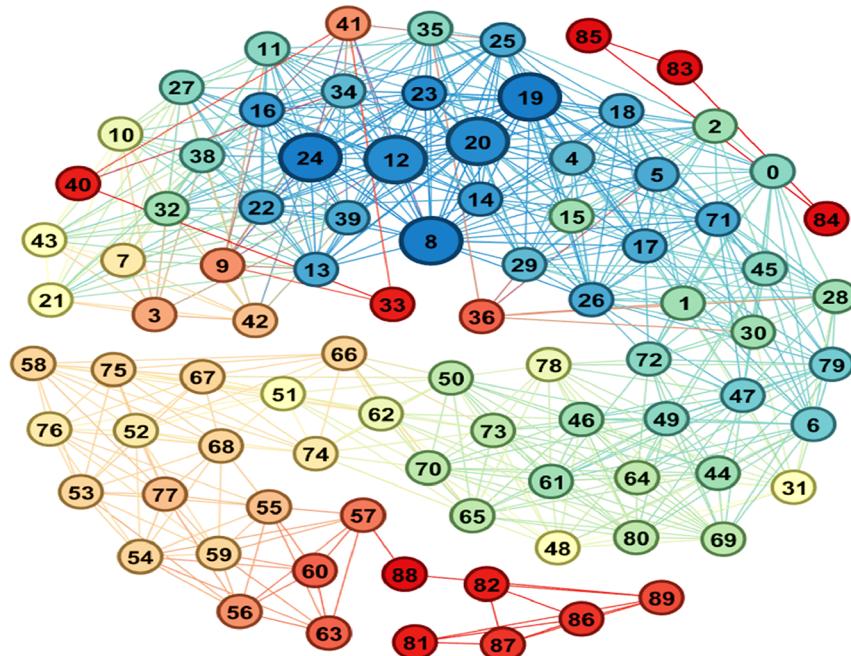
$$\text{sig}(V_{ij}) = \left| \frac{1 - \exp(-V_{ij})}{1 + \exp(-V_{ij})} \right| \quad (12)$$

where  $k = \text{rand} \times \deg(n_j)$ ,  $k \neq X_{ij}(t)$ ,  $\deg(n_j)$  is the degree of the vertex found in  $n_j$ , and  $\rho$  is the threshold set by the user.

Another noteworthy point to mention involves the generated positional value based on the distribution degree, which indicates that if the value of node  $v_j$  is greater than those of its surrounding neighbors or if  $\text{sig}(V_{ij}(t+1))$  is ever greater than the value of  $\rho$ , the neighbors of the nodes must then be transferred to the currently selected neighbors. Therefore, the sigmoid function  $\text{sig}()$  function found within eq 11 is modified to solve this issue. The particle position is very likely to change



**Figure 3.** Degree distribution of the HEA interaction network. The degree distribution presents the probability distribution of degrees over the whole network. The degree of each compound is the number of edges that it has to other compounds.



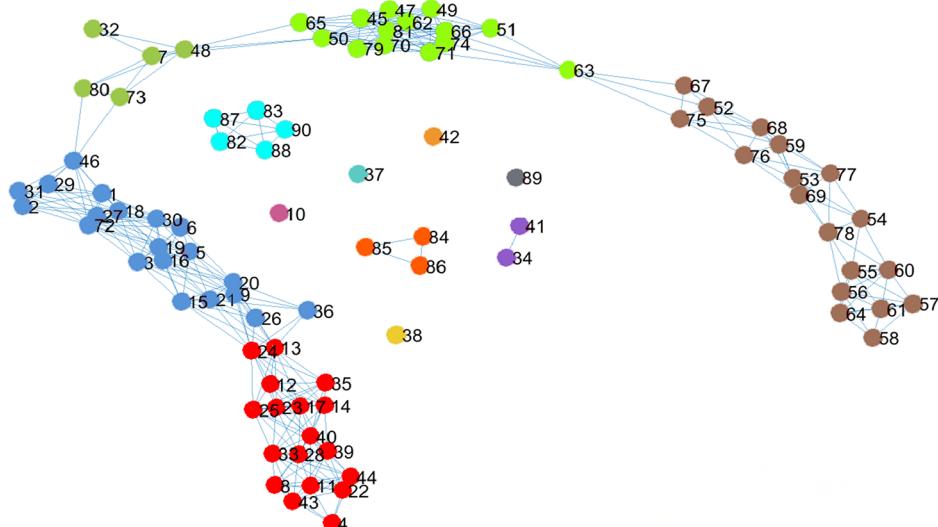
**Figure 4.** Impact of threshold and  $\alpha$  on the HEA interaction network. The threshold prunes the weak connection in the HEA network. The  $\alpha$  coefficient determines the value of content and structural similarity.

through the particle velocity reduction procedure, causing the PSO to gradually converge in the global optimization.<sup>23</sup>

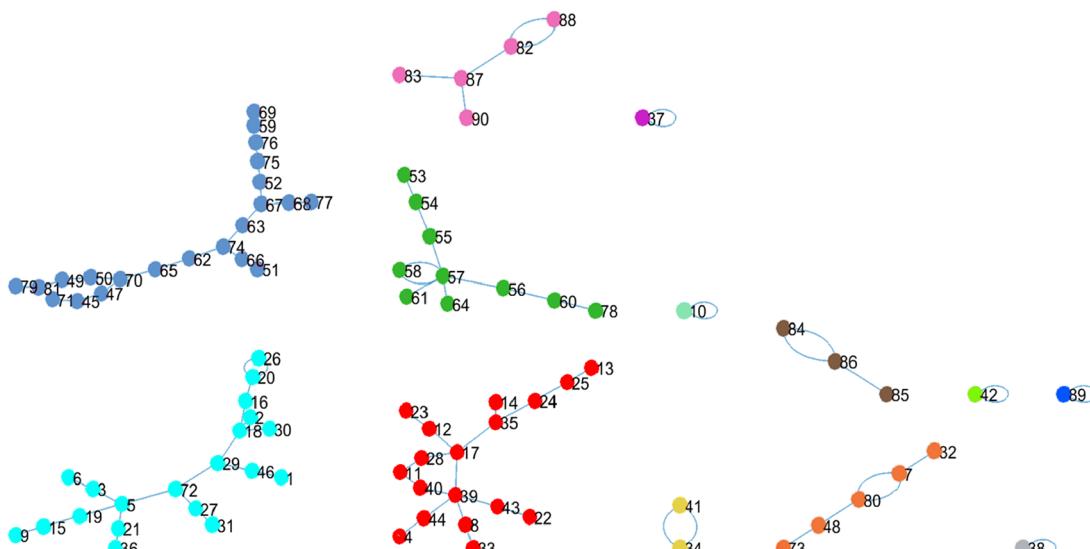
## 5. RESULTS AND DISCUSSION

In view of the novelty of this study, several experiments were carried out to evaluate the efficiency of the proposed method. In order to detect the HEA interaction network, the similarity of alloy features must be addressed. The weighted interaction network of the HEAs is shown in Figure 2, where the weight of the links between the compounds determines the degree of similarity among them. This particular interaction network was

initially formed based on the content and structural similarity of the alloy descriptors. Besides, all compositions were linked leading to the formation of a complete graph. The primary interaction network had 90 sets of nodes corresponding to the HEA compounds, and each compound is shown by the number presented in results as defined in Appendix A. As shown in Figure 2, every compound was linked to every other compound using 3968 edges. The interaction network is an undirected graph where all compounds are connected to each other. The degree of each node in the HEA interaction network is the number of edges that it has to other nodes. The



**Figure 5.** Community detection with the use of the developed Louvain algorithm. Any colored community shows a community and the compounds have similar functionality and descriptors.



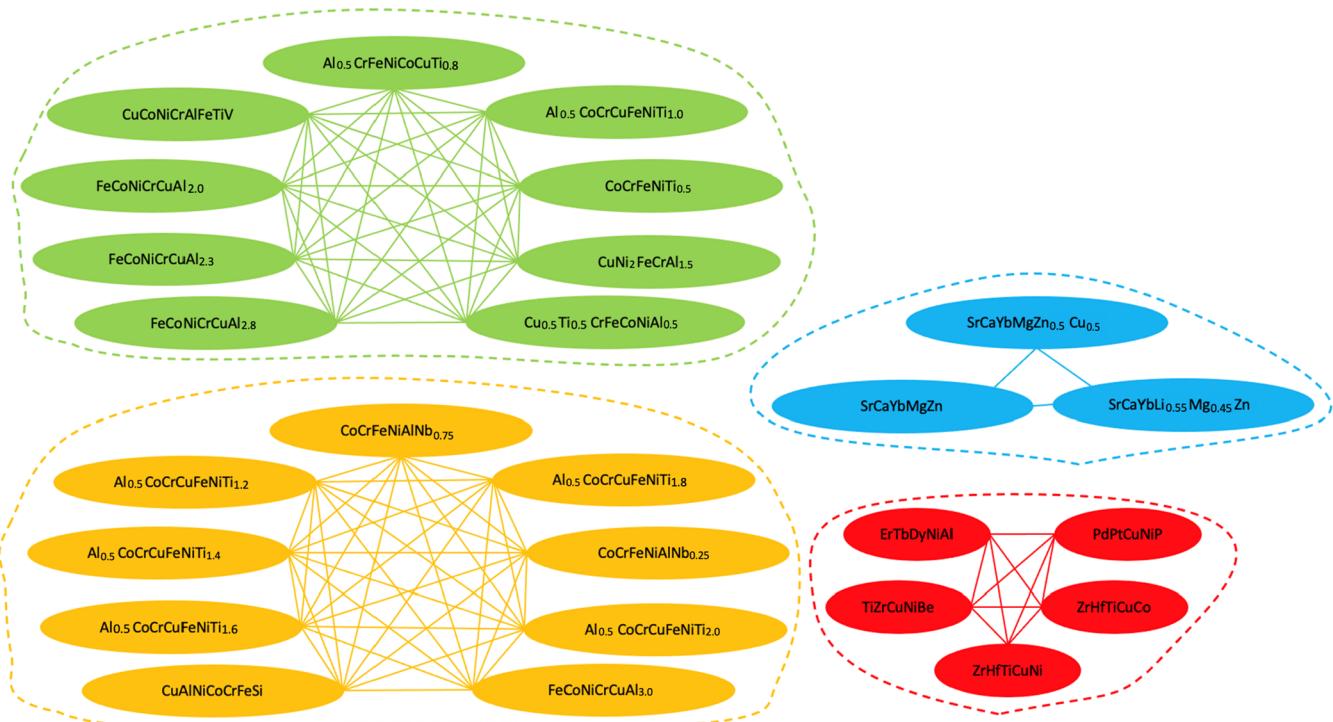
**Figure 6.** Community detection with the use of an optimal PSO algorithm. Any colored community presents a community of HEA compound that is extracted by the PSO algorithm.

degree distribution shows the probability distribution of degrees over the whole network. The degree distribution diagram for the graph constructed is shown in Figure 3. As shown in Figure 3, the average degree based on the diagram is 14.20, which is the probability distribution of these degrees over the network. The next step is illustrated in Figure 4, an  $\alpha$  coefficient value of 0.9 and a threshold of 0.6 are both applied to the network, which eliminates less similar nodes and weak connections. As shown in Figure 4, the resulting interaction network contains 632 edges to maintain the communications among all 90 nodes. In Figure 4, the nodes are drawn with the size reflecting the degree of each node.

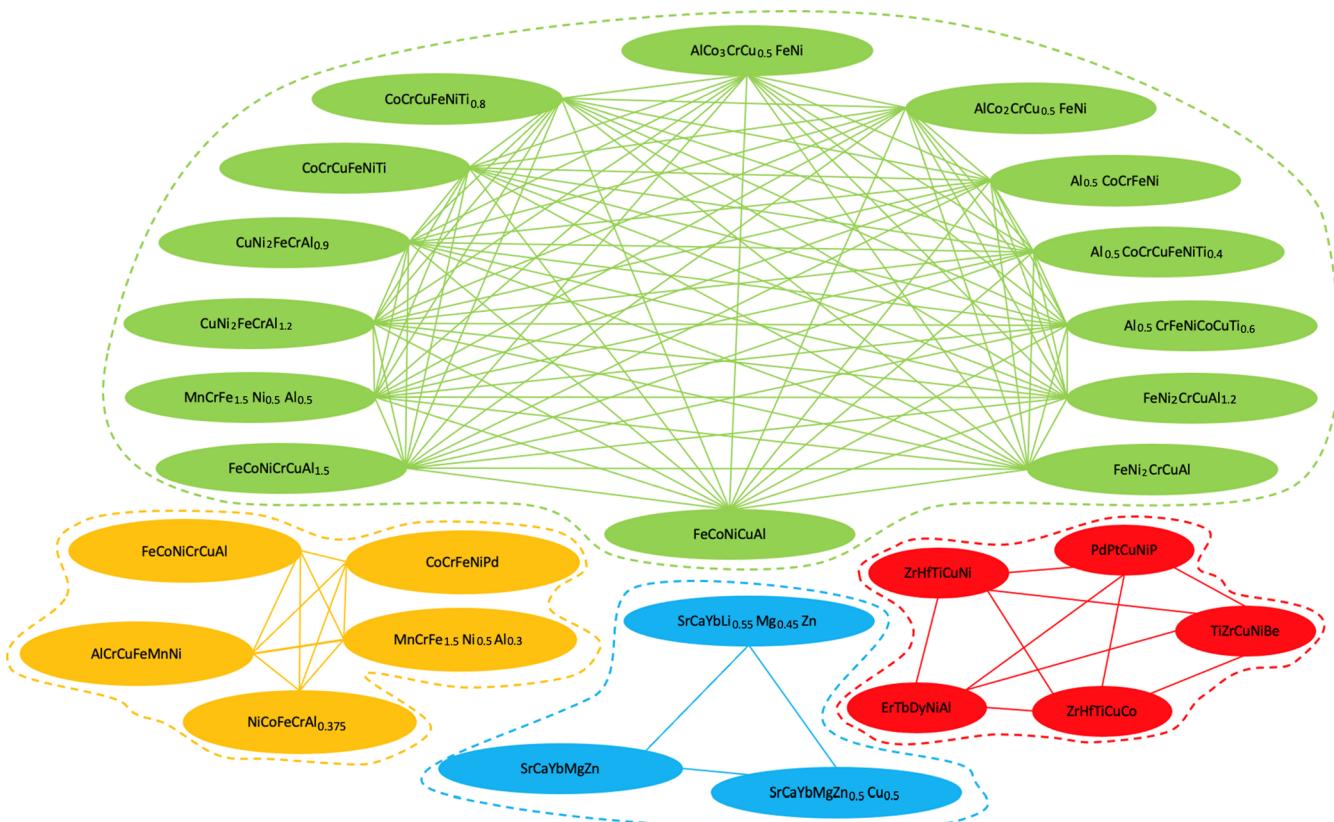
As presented in Figure 5, the Louvain algorithm has been applied to the HEA interaction network, which extracts 13 communities with an overall quality of approximately 0.71. For these 13 communities shown in Figure 5, each community is indicated by a unique color, and the compounds in every community are fully connected. As shown in Figure 6, the communities are also extracted with the optimal PSO

algorithm using the HEA interaction network in 100 iterations. The 13 optimal communities by PSO are displayed with an enhanced quality of approximately 0.89 shown in Figure 6. Because the compositions of each community are not connected to the other communities' compounds, the communities obtained from the optimal PSO algorithm have a higher quality modularity parameter. The analysis of each community shows that the neighbors of every compound have the same phase label and include similar elements.

In this paper, the measurement criteria for both community recognition algorithms are the main parameters for assessing the quality of communities.<sup>14</sup> If the number of edges found in a community is not more than a random diagram, it can be concluded that the modularity is zero. Another point to note is the maximum modular value, which is basically obtained when all the internal nodes within a community are connected and there is no external edge to other communities.<sup>14</sup> One of the basic features of modularity is the ability to compare different communities with various methods. Because other algorithms



**Figure 7.** Four community samples obtained through the developed Louvain algorithm. These four communities are selected as an example from the results that present the functionality of the proposed method.



**Figure 8.** Four community samples obtained through the PSO method. These four communities are selected as an example from the results that present the functionality of the proposed method.

do not necessarily extract the same results, many existing criteria cannot assess the quality of communities. Therefore, using the Louvain's hierarchical bottom-up method, analysis of

modularity trends in the process of dividing or merging communities can be investigated. The maximum value of this parameter is considered as the best outcome. Moreover, the

**Table 2.** Precision of Louvain and PSO Algorithms in the Phase Prediction of HEA Alloys Indicates That Communities Can Improve the Phase Prediction Precision

community number	compound phase by PSO	community prediction phase by PSO	precision by PSO %	compound phase by Louvain	community prediction phase by Louvain	precision by Louvain %
1	38	30	78	19	18	94
2	19	17	89.5	18	13	2
3	1	1	100	5	2	40
4	2	1	50	1	1	100
5	1	1	100	2	1	50
6	1	1	100	1	1	100
7	1	1	100	1	1	100
8	9	9	100	1	1	100
9	9	9	100	14	14	100
10	5	5	100	18	18	100
11	3	3	100	5	5	100
12	1	1	100	3	3	100
total average precision		93.125		total average precision		88.019

modularity of each community is a scalar value between  $-1$  and  $1$ , which essentially measures the density of the community's internal links in comparison with the links found between communities;<sup>15,16</sup> a modularity between  $0.3$  and  $0.7$  indicates a strong community.<sup>18</sup> Because this criterion is closer to  $1$ , the communities are high quality. According to the experimental result, the modularity of both optimal algorithms in this paper is upper than  $0.7$  (Table 1).

The four communities are, as an example, shown in Figure 7, for communities extracted based on the developed Louvain algorithm. For example, in the blue community (Figure 7), SrCaYbMgZn, SrCaYbMgZn<sub>0.5</sub>Cu<sub>0.5</sub>, and SrCaYbLi<sub>0.55</sub>Mg<sub>0.45</sub>Zn are in a single community with similar elements including Sr, Ca, Yb, Mg, and Zn which are found in these alloys with equal functionality. The number of elements in the blue community (Figure 7) is five and six, where two alloys SrCaYbMgZn<sub>0.5</sub>Cu<sub>0.5</sub> and SrCaYbLi<sub>0.55</sub>Mg<sub>0.45</sub>Zn have the same number of elements and contain common elements. The alloys in the blue community (Figure 7) are amorphous. The descriptor values for the alloys in the blue community (Figure 7) vary in ranges:  $\delta$  [15.25, 16.37],  $\Delta H_{\text{mix}}$  [-10.6, -13.12],  $S_c$  [1.61, 1.75],  $\varphi$  [-0.017, 0.61],  $\epsilon_{\text{RMS}}$  [0.1565, 0.1699], and VEC [4.09, 4.2]. The yellow community (Figure 7) included nine HEA compounds, such as CoCrFeNiAlNb<sub>0.75</sub>, Al<sub>1.0</sub>.<sub>5</sub>CoCrCuFeNiTi<sub>1.2</sub>, Al<sub>0.5</sub>CoCrCuFeNiTi<sub>1.4</sub>, Al<sub>0.5</sub>CoCrCuFeNiTi<sub>1.6</sub>, CuAlNiCoCr-FeSi, Al<sub>0.5</sub>CoCrCuFeNiTi<sub>1.8</sub>, CoCrFeNiAlNb<sub>0.25</sub>, Al<sub>0.5</sub>CoCrCuFeNiTi<sub>2.0</sub>, and FeCoNiCrCuAl<sub>3.0</sub>. They have five common elements including Al, Co, Cr, Fe, and Ni. The HEA compounds in the yellow community are multiphase alloys. They contain six and seven elements and their descriptors have a similar range for three compounds. The  $\delta$  parameter is in the range [6.1, 7.31],  $\Delta H_{\text{mix}}$  in the range [-18.86, -10.56],  $S_c$  in the range [1.67, 1.95],  $\varphi$  in the range [3.91, 6.26],  $\epsilon_{\text{RMS}}$  in the range [0.0605, 0.0729], and VEC in the range [6.63, 7.51]. The green community included nine HEA compounds, such as Al<sub>0.5</sub>CrFeNiCoCuTi<sub>0.8</sub>, CuCoNiCrAlFeTiV, FeCoNiCr-CuAl<sub>2.0</sub>, FeCoNiCrCuAl<sub>2.3</sub>, FeCoNiCrCuAl<sub>2.8</sub>, Al<sub>0.5</sub>CoCrCuFeNiTi<sub>1.0</sub>, CoCrFeNiTi<sub>0.5</sub>, CuNi<sub>2</sub>FeCrAl<sub>1.5</sub>, and Cu<sub>0.5</sub>Ti<sub>0.5</sub>CrFeCoNiAl<sub>0.5</sub>. They have three common elements including Cr, Fe, and Ni. The HEA compounds in the green community are multiphase alloys. The  $\delta$  parameter is in the range [5.33, 6.57],  $\Delta H_{\text{mix}}$  in the range [-13.94, -8.05],  $S_c$  in the range [1.57, 2.08],  $\varphi$  in the range [5.53, 8.79],  $\epsilon_{\text{RMS}}$  in the

range [0.0525, 0.0656], and VEC in the range [6.71, 7.78]. The red community (Figure 7) included PdPtCuNiP, TiZrCuNiBe, ZrHfTiCuCo, ErTbDyNiAl, and ZrHfTiCuNi alloys. All alloys within a red community consist of five elements. Furthermore, ZrHfTiCuNi and ZrHfTiCuCo have common elements such as Zr, Hf, Ti, and Cu. All the alloys within that community are amorphous. The descriptor values for alloys in the red community are in the range  $\delta$  [9.29, 13.74],  $\Delta H_{\text{mix}}$  [-23.52, -37.6],  $\varphi$  [-2.24, 0.42],  $\epsilon_{\text{RMS}}$  [0.0952, 0.1429], VEC [4.4, 9.2], and  $S_c$  is 1.61, equal for all alloys in the red cluster.

The four communities are illustrated in Figure 8, which is obtained based on the PSO algorithm. As an example, the yellow community included FeCoNiCrCuAl, AlCrCuFeMnNi, NiCoFeCrAl<sub>0.375</sub>, MnCrFe<sub>1.5</sub>Ni<sub>0.5</sub>Al<sub>0.3</sub>, and CoCrFeNiPd compounds. The common elements in a community are Fe, Ni, and Cr. The alloys in the yellow community (Figure 8) contain five and six elements. Two alloys FeCoNiCrCuAl and AlCrCuFeMnNi have the same number of elements and contain five common elements. Some alloys in the yellow community (Figure 8) are multiphase, and some of them are single phase (FCC). The descriptor values for the alloys in the yellow community (Figure 8) vary  $\delta$  in the range [4.12, 5.39],  $\Delta H_{\text{mix}}$  in the range [-7.99, -4.78],  $S_c$  in the range [1.48, 1.79],  $\varphi$  in the range [13.54, 16.16],  $\epsilon_{\text{RMS}}$  in the range [0.0406, 0.0536], and VEC in the range [7.19, 8.8]. The blue community included SrCaYbLi<sub>0.55</sub>Mg<sub>0.45</sub>Zn, SrCaYbMgZn, and SrCaYbMgZn<sub>0.5</sub>Cu<sub>0.5</sub>, which is the same as in the blue community (Figure 7) extracted by the Louvain method. The green community (Figure 8) included 14 HEA compounds, such as AlCo<sub>3</sub>CrCu<sub>0.5</sub>FeNi, CoCrCuFeNiTi<sub>0.8</sub>, CoCrCuFeNi-Ti, CuNi<sub>2</sub>FeCrAl<sub>0.9</sub>, CuNi<sub>2</sub>FeCrAl<sub>1.2</sub>, MnCrFe<sub>1.5</sub>Ni<sub>0.5</sub>Al<sub>0.5</sub>, FeCoNiCrCuAl<sub>1.5</sub>, FeCoNiCuAl, FeNi<sub>2</sub>CrCuAl, Fe-Ni<sub>2</sub>CrCuAl<sub>1.2</sub>, Al<sub>0.5</sub>CrFeNiCoCuTi<sub>0.6</sub>, Al<sub>0.5</sub>CoCrCuFeNiTi<sub>0.4</sub>, Al<sub>0.5</sub>CoCrFeNi, and AlCo<sub>2</sub>CrCu<sub>0.5</sub>FeNi. They have common elements, such as Fe and Ni. The HEAs contain five and six elements. All the HEA compounds in the green community (Figure 8) are multiphase. The  $\delta$  parameter is in the range [4.6, 6.12],  $\Delta H_{\text{mix}}$  in the range [-9.09, -5.22],  $S_c$  in the range [1.52, 1.92],  $\varphi$  in the range [8.92, 13.02],  $\epsilon_{\text{RMS}}$  in the range [0.0454, 0.0605], and VEC in the range [7.0, 8.2]. The red community (Figure 8) included PdPtCuNiP, TiZrCuNiBe, ZrHfTiCuCo, ErTbDyNiAl, and ZrHfTiCuNi alloys, which is

**Table A1. Precision of Louvain and PSO Algorithms in the Phase Prediction of HEA Alloys Indicates That Communities Can Improve the Phase Prediction Precision**

compound number	HEA alloy composition	$\delta$ (%)	$\Delta H_{\text{mix}}$ (kJ/mol)	$S_c$ ( $k_B$ /atom)	$\varphi$	$\epsilon_{\text{RMS}}$	VEC
1	Al <sub>0.5</sub> CoCrCuFeNiTi <sub>0.2</sub>	4.93	-4.15	1.86	17.41	0.0487	8.12
2	Al <sub>0.3</sub> CoCrFeNi	3.76	-7.27	1.54	19.99	0.037	7.88
3	Al <sub>0.5</sub> CrCuFeNi <sub>2</sub>	4.2	-2.51	1.52	20.44	0.0414	8.45
4	CoCrFeNi	0.3	-3.75	1.39	3583.31	0.0039	8.25
5	CoFeMnNi	3.55	-4	1.39	24.51	0.0353	8.5
6	CoCrMnNi	3.45	-5.5	1.39	23.99	0.0343	8
7	CoCrFeNiPd	4.46	-7.04	1.61	15.95	0.0446	8.8
8	CoCrCu <sub>0.5</sub> FeNi	0.84	0.49	1.58	627.5	0.0083	8.56
9	CuNiCoFeCrAl <sub>0.5</sub> V <sub>0.2</sub>	4.15	-2.5	1.86	26.49	0.0409	8.16
10	CuNiFeCrMo	3.58	4.64	1.61	28.97	0.0356	8.2
11	CuNiCoFe	1.14	5	1.39	223.55	0.0114	9.5
12	CuNiCoFeMn	3.18	1.76	1.61	41.04	0.0316	9
13	CuNi <sub>2</sub> FeMn <sub>2</sub> Cr	3.57	-0.49	1.55	33.58	0.0356	8.43
14	CuNi <sub>2</sub> FeCrAl <sub>0.2</sub>	2.94	0.12	1.44	44.59	0.0289	8.77
15	CuNi <sub>2</sub> FeCrAl <sub>0.4</sub>	3.86	-1.7	1.5	24.87	0.0381	8.56
16	CuNi <sub>2</sub> FeCrAl <sub>0.5</sub>	4.2	-2.51	1.52	20.44	0.0414	8.45
17	Cu <sub>0.75</sub> NiCoFeCrAl <sub>0.25</sub>	3.25	-0.71	1.72	42.39	0.032	8.4
18	Cu <sub>0.5</sub> NiCoFeAl <sub>0.5</sub> Cr	4.37	-4.6	1.75	20.16	0.0431	8
19	Cu <sub>0.5</sub> NiCoCrAl <sub>0.5</sub> Fe <sub>2</sub>	4.08	-3.53	1.68	23.06	0.0403	8
20	Cu <sub>0.5</sub> NiCoCrAl <sub>0.5</sub> Fe <sub>3</sub>	3.84	-2.84	1.57	24.99	0.0379	8
21	Cu <sub>0.5</sub> NiCoCrAl <sub>0.5</sub> Fe <sub>3.5</sub>	3.74	-2.58	1.52	25.76	0.0368	8
22	FeCoNiCrCu	1.03	3.2	1.61	369.34	0.0103	8.8
23	FeNi <sub>2</sub> CrCuAl <sub>0.2</sub>	2.94	0.12	1.44	44.59	0.0289	8.77
24	FeCrMnNiCo	3.27	-4.16	1.61	34.71	0.0325	8
25	FeCoNiCrCuAl <sub>0.3</sub>	3.42	0.16	1.79	44.96	0.0337	8.47
26	FeCoNiCrCuAl <sub>0.5</sub>	4.17	-1.52	1.77	25.77	0.0411	8.27
27	FeNi <sub>2</sub> CrCuAl <sub>0.6</sub>	4.49	-3.27	1.53	17.39	0.0443	8.36
28	NiCoFeCrMo <sub>0.3</sub>	2.38	-4.15	1.54	62.1	0.0235	8.09
29	NiCoFeCrMo <sub>0.1</sub> Al <sub>0.3</sub>	3.9	-7.26	1.62	20.05	0.0385	7.84
30	NiCoFeCrAl <sub>0.25</sub>	3.48	-6.75	1.53	23.78	0.0342	7.94
31	NiCoFeCrAl <sub>0.3</sub>	3.76	-7.27	1.54	19.99	0.037	7.88
32	NiCoFeCrAl <sub>0.375</sub>	4.12	-7.99	1.56	16.16	0.0406	7.8
33	VCuFeCoNi	2.2	-2.24	1.61	84.95	0.022	8.6
34	TaNbHfZrTi	4.99	2.72	1.61	16.9	0.0499	4.4
35	TaNbVTi	3.93	-0.25	1.39	26.07	0.0397	4.75
36	TaNbVTiAl <sub>0.25</sub>	3.83	-4.82	1.53	25.84	0.0387	4.65
37	TaNbVTiAl <sub>0.5</sub>	3.74	-8.4	1.58	24.28	0.0377	4.56
38	TaNbVTiAl <sub>1.0</sub>	3.57	-13.44	1.61	20.38	0.036	4.4
39	WNbMoTa	2.31	-6.5	1.39	60.87	0.0231	5.5
40	WNbMoTaV	3.15	-4.64	1.61	41.18	0.0315	5.4
41	Al <sub>2</sub> Li <sub>20</sub> Mg <sub>10</sub> Sc <sub>20</sub> Ti <sub>30</sub>	5.16	-0.4	1.56	16.17	0.0515	2.8
42	GdTbDyTmLu	5.07	0	1.61	18.76	0.0515	3
43	HoDyYGdTb	0.81	0	1.61	701.52	0.0081	3
44	YgdTbDyLu	1.37	0	1.61	245.87	0.0137	3
45	AlCo <sub>3</sub> CrCu <sub>0.5</sub> FeNi	4.88	-7.25	1.62	12.52	0.0482	7.93
46	Al <sub>0.8</sub> CrCuFeMnNi	5.15	-3.97	1.79	15.73	0.0512	7.66
47	AlCo <sub>2</sub> CrCu <sub>0.5</sub> FeNi	5.17	-7.67	1.71	11.83	0.0511	7.77
48	AlCrCuFeMnNi	5.39	-5.11	1.79	13.54	0.0536	7.5
49	Al <sub>0.5</sub> CoCrFeNi	4.6	-9.09	1.58	12.23	0.0454	7.67
50	Al <sub>0.5</sub> CoCrCuFeNiTi <sub>0.4</sub>	5.49	-6.42	1.9	13.02	0.0543	7.98
51	Al <sub>0.5</sub> CrFeNiCoCuTi <sub>0.6</sub>	5.92	-8.4	1.92	10.36	0.0586	7.85
52	Al <sub>0.5</sub> CrFeNiCoCuTi <sub>0.8</sub>	6.26	-10.11	1.92	8.54	0.0621	7.73
53	Al <sub>0.5</sub> CoCrCuFeNiTi <sub>1.0</sub>	6.53	-11.6	1.93	7.23	0.0649	7.62
54	Al <sub>0.5</sub> CoCrCuFeNiTi <sub>1.2</sub>	6.76	-12.89	1.92	6.26	0.0671	7.51
55	Al <sub>0.5</sub> CoCrCuFeNiTi <sub>1.4</sub>	6.94	-14.02	1.91	5.47	0.069	7.41
56	Al <sub>0.5</sub> CoCrCuFeNiTi <sub>1.6</sub>	7.09	-15.01	1.9	4.85	0.0706	7.31
57	Al <sub>0.5</sub> CoCrCuFeNiTi <sub>1.8</sub>	7.21	-15.86	1.89	4.34	0.0719	7.22
58	Al <sub>0.5</sub> CoCrCuFeNiTi <sub>2.0</sub>	7.31	-16.6	1.88	3.91	0.0729	7.13
59	CoCrFeNiTi <sub>0.5</sub>	5.33	-11.56	1.58	7.91	0.0525	7.78
60	CoCrFeNiAlNb <sub>0.25</sub>	6.1	-14.66	1.72	5.26	0.0605	7.1

**Table A1.** continued

compound number	HEA alloy composition	$\delta$ (%)	$\Delta H_{\text{mix}}$ (kJ/mol)	$S_c$ ( $k_B$ /atom)	$\varphi$	$\varepsilon_{\text{RMS}}$	VEC
61	CoCrFeNiAlNb <sub>0.75</sub>	6.5	-18.03	1.79	3.95	0.0648	6.91
62	CoCrCuFeNiTi <sub>0.8</sub>	5.7	-6.75	1.79	11.12	0.0563	8.14
63	CoCrCuFeNiTi	6.12	-8.44	1.79	8.92	0.0605	8
64	CuAlNiCoCrFeSi	6.13	-18.86	1.95	4.15	0.061	7.29
65	CuNi <sub>2</sub> FeCrAl <sub>0.9</sub>	5.15	-5.22	1.56	12.08	0.0509	8.08
66	CuNi <sub>2</sub> FeCrAl <sub>1.2</sub>	5.6	-6.78	1.57	9.25	0.0556	7.83
67	CuNi <sub>2</sub> FeCrAl <sub>1.5</sub>	5.93	-8.05	1.57	7.47	0.0589	7.62
68	Cu <sub>0.5</sub> Ti <sub>0.5</sub> CrFeCoNiAl <sub>0.5</sub>	5.97	-10.84	1.89	8.79	0.0591	7.64
69	CuCoNiCrAlFeTiV	6.34	-13.94	2.08	7.73	0.0631	7
70	FeNi <sub>2</sub> CrCuAl	5.32	-5.78	1.56	10.94	0.0526	8
71	FeNi <sub>2</sub> CrCuAl <sub>1.2</sub>	5.6	-6.78	1.57	9.25	0.0555	7.84
72	FeCoNiCrCuAl <sub>0.8</sub>	4.92	-3.61	1.79	17.15	0.0487	8
73	FeCoNiCrCuAl	5.28	-4.78	1.79	14.12	0.0523	7.83
74	FeCoNiCrCuAl <sub>1.5</sub>	5.89	-7.05	1.78	9.9	0.0585	7.46
75	FeCoNiCrCuAl <sub>2.0</sub>	6.26	-8.65	1.75	7.62	0.0623	7.14
76	FeCoNiCrCuAl <sub>2.3</sub>	6.4	-9.38	1.73	6.7	0.0638	6.97
77	FeCoNiCrCuAl <sub>2.8</sub>	6.57	-10.28	1.68	5.53	0.0656	6.71
78	FeCoNiCrCuAl <sub>3.0</sub>	6.61	-10.56	1.67	5.17	0.0661	6.63
79	FeCoNiCuAl	5.61	-5.28	1.61	10.44	0.0556	8.2
80	MnCrFe <sub>1.5</sub> Ni <sub>0.5</sub> Al <sub>0.3</sub>	4.7	-5.51	1.48	13.89	0.047	7.19
81	MnCrFe <sub>1.5</sub> Ni <sub>0.5</sub> Al <sub>0.5</sub>	5.16	-7.26	1.52	10.62	0.051	7
82	ErTbDyNiAl	13.74	-37.6	1.61	-2.24	0.1429	4.4
83	PdPtCuNiP	9.29	-23.68	1.61	-1.26	0.0952	9.2
84	SrCaYbMgZn	15.25	-13.12	1.61	-0.017	0.1565	4.2
85	SrCaYbMgZn <sub>0.5</sub> Cu <sub>0.5</sub>	16.37	-10.6	1.75	0.61	0.1699	4.1
86	SrCaYbLi <sub>0.55</sub> Mg <sub>0.45</sub> Zn	15.71	-12.15	1.75	0.2	0.1612	4.09
87	TiZrCuNiBe	12.53	-30.24	1.61	-0.9	0.1268	6.2
88	ZrHfTiCuNi	10.34	-27.36	1.61	-0.27	0.1049	6.6
89	ZrHfTiCuFe	10.43	-15.84	1.61	1.73	0.1059	6.2
90	ZrHfTiCuCo	10.24	-23.52	1.61	0.42	0.1039	6.4

the same as the red community (Figure 7) extracted by the Louvain method.

Considering the communities in the present article, it can be concluded that these clusters essentially have high quality and accuracy, which can be shown through the modularity criterion. Table 1 shows the obtained results of the modularity criterion of community quality using the developed Louvain algorithm and optimal PSO method. As is shown in Table 1, the modularity criteria in the developed Louvain algorithm is a constant value equal to 0.71, and it has not changed. Also, in optimal PSO, the modularity parameter is started from 0.87 and after 30 iterations, it increased up to 0.89 and after 60 iterations until 150 iterations, it did not change any more. Therefore, the modularity value by optimal PSO is about 0.89. Finally, the benefits of HEA community detection are discussed. In the field of biology and proteins, the analysis of protein networks is useful because proteins that are in a community have common behaviors and properties. This means that proteins of the same community behave similarly. Therefore, it can be concluded that the purpose of community detection is to have HEAs with similar elements in their composition and phases, such as colored community alloys (Figure 7). Considering the fact that HEAs in a community have similar properties, one can recognize the properties of HEAs as soon as the community of alloys is determined. For example, the maximal number of elements in the alloy could be predicted according to the communities extracted.

Another advantage of community detection in HEA interaction networks is the phase prediction using ML

techniques, which is presented in this paper. Because the most extracted communities have the same phase in the HEA network, the phase composition of each compound, which is indistinct, can be anticipated. The unseen compound's phase can be identified by the other compounds that are in the same community. As shown in Table 2, we looked at the number of phases in the same community. In addition, the precision of phase prediction is shown in Table 2 showing the phase forecast by Leuven and PSO, which is approximately 88% and approximately 93%, respectively.

## 6. DATA AND SOFTWARE AVAILABILITY

In this study, we have used MATLAB software version R2019a, which has been referred in <https://www.mathworks.com/>. As described and referenced in Section 3, the data set and source code for this paper are located at the GitHub. The data set referred <https://github.com/rghoochannejad/HEAs-Community-Detection/tree/Dataset> and the source code is referred at <https://github.com/rghoochannejad/HEAs-Community-Detection/tree/main>. The data set to this article can be found online at <https://doi.org/10.1016/j.mattod.2015.11.026>.

## 7. CONCLUSIONS AND SUGGESTIONS FOR FUTURE RESEARCH

The present study aimed to present a novelty for community detection based on ML to detect HEA compounds that behave similarly to each other. At first, the descriptors of each compound are analyzed, and then the similarities among the

alloys in terms of phase composition are calculated accordingly. Second, an interaction network of HEAs is established, which could very well be linked to the interaction network. Additionally, both the quality and accuracy of extractive communities and their modularity criteria have been analyzed and investigated thoroughly using two methods of Louvain and PSO algorithms, indicating that the proposed method has a high quality in community detection. This evaluation shows that the detected clusters potentially have robust internal connections among the compounds. Although the obtained results of the current method were indicative of high quality and precision, it does not mean that it cannot be further developed. It is also important to mention that other methods can be implemented very well in future studies to determine the more advanced properties of alloys. The present method can also be developed in larger data sets with maintaining the quality. The use of other ML methods still have great potential for obtaining better results, although these statistical methods and ML algorithms do in fact enhance the speed of the research conduction in the field of materials science. The introduced method is not considered as the only efficient way for community detection, but it can be applied in other areas of materials science leading to the detection of other beneficial alloy compositions that can be used in the industry. Finally, the HEA community detection is useful to finding new common features of similar alloys. Moreover, phase prediction is an action, which can be performed by community detection in this study with a good precision rate.

## APPENDIX A

The HEA alloy compositions and descriptors of HEAs are given in Table A1.<sup>1</sup>

## AUTHOR INFORMATION

### Corresponding Author

Mehrdad Jalali – Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran; Institute of Functional Interfaces (IFG), Karlsruhe Institute of Technology (KIT), 76344 Eggenstein-Leopoldshafen, Germany; [orcid.org/0000-0003-2465-4933](https://orcid.org/0000-0003-2465-4933); Email: [Mehrdad.jalali@kit.edu](mailto:Mehrdad.jalali@kit.edu)

### Authors

Raheleh Ghouchan Nezhad Noor Nia – Department of Computer Engineering, Mashhad Branch, Islamic Azad University, Mashhad, Iran; [orcid.org/0000-0001-5697-9041](https://orcid.org/0000-0001-5697-9041)

Matthias Mail – Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), 76344 Eggenstein-Leopoldshafen, Germany; Karlsruhe Nano Micro Facility (KNMF), 76344 Eggenstein-Leopoldshafen, Germany

Yulia Ivanisenko – Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), 76344 Eggenstein-Leopoldshafen, Germany

Christian Kübel – Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), 76344 Eggenstein-Leopoldshafen, Germany; Karlsruhe Nano Micro Facility (KNMF), 76344 Eggenstein-Leopoldshafen, Germany; Department of Materials & Geological Sciences, Technical University Darmstadt, 64287 Darmstadt, Germany; [orcid.org/0000-0001-5701-4006](https://orcid.org/0000-0001-5701-4006)

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acsomegab00317>

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors acknowledge funding by the Federal Ministry of Education and Research of Germany for the project STREAM (“Semantische Repräsentation, Vernetzung und Kuratierung von qualitätsgesicherten Materialdaten”, ID: 16QK11C). Furthermore, the authors acknowledge support by the KIT-Publication Fund of the Karlsruhe Institute of Technology.

## REFERENCES

- (1) Ye, Y. F.; Wang, Q.; Lu, J.; Liu, C. T.; Yang, Y. High-entropy alloy: challenges and prospects. *Mater. Today* **2016**, *19*, 349–362.
- (2) Cantor, B.; Chang, I. T. H.; Knight, P.; Vincent, A. J. B. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng., A* **2004**, *375–377*, 213–218.
- (3) Yeh, J.-W.; Chen, S.-K.; Lin, S.-J.; Gan, J.-Y.; Chin, T.-S.; Shun, T.-T.; Tsau, C.-H.; Chang, S.-Y. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **2004**, *6*, 299–303.
- (4) Chou, H.-P.; Chang, Y.-S.; Chen, S.-K.; Yeh, J.-W. Microstructure, thermophysical and electrical properties in  $\text{Al}_x\text{CoCrFeNi}$  ( $0 \leq x \leq 2$ ) high-entropy alloys. *Mater. Sci. Eng., B* **2009**, *163*, 184–189.
- (5) Chen, Y. Y.; Hong, U. T.; Shih, H. C.; Yeh, J. W.; Duval, T. Electrochemical kinetics of the high entropy alloys in aqueous environments—a comparison with type 304 stainless steel. *Corros. Sci.* **2005**, *47*, 2679–2699.
- (6) Ye, Y. F.; Wang, Q.; Lu, J.; Liu, C. T.; Yang, Y. Design of high entropy alloys: A single-parameter thermodynamic rule. *Scr. Mater.* **2015**, *104*, 53–55.
- (7) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555.
- (8) Dai, D.; Xu, T.; Wei, X.; Ding, G.; Xu, Y.; Zhang, J.; Zhang, H. Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys. *Comput. Mater. Sci.* **2020**, *175*, 109618.
- (9) Zhang, Y.; Zhou, Y. J.; Lin, J. P.; Chen, G. L.; Liaw, P. K. Solid-solution phase formation rules for multi-component alloys. *Adv. Eng. Mater.* **2008**, *10*, 534–538.
- (10) He, Q.; Yang, Y. On lattice distortion in high entropy alloys. *Front. Mater.* **2018**, *5*, 42.
- (11) Xia, P.; Zhang, L.; Li, F. Learning similarity with cosine similarity ensemble. *Inf. Sci.* **2015**, *307*, 39–52.
- (12) Niwattanakul, S.; Singthongchai, J.; Naenudorn, E.; Wanapu, S. Using of Jaccard coefficient for keywords similarity. *Proceedings of the International Multiconference of Engineers and Computer Scientists*, 2013; pp 380–384.
- (13) Jain, S.; Shukla, S.; Wadhvani, R. Dynamic selection of normalization techniques using data complexity measures. *Expert Syst. Appl.* **2018**, *106*, 252–262.
- (14) Newman, M. E. J. Fast algorithm for detecting community structure in networks. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2004**, *69*, 066133.
- (15) Girvan, M.; Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7821–7826.
- (16) Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8577–8582.
- (17) Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theor. Exp.* **2008**, *2008*, P10008.
- (18) Ponciano, J. R.; Linhares, C. D. G.; Melo, S. L.; Lima, L. V.; Travençolo, B. A. N. Visual analysis of contact patterns in school environments. *Inf. Educ.* **2020**, *19*, 455–472.

- (19) Shirjini, M. F.; Farzi, S.; Nikanjam, A. MDPCluster: a swarm-based community detection algorithm in large-scale graphs. *Computing* **2020**, *102*, 893.
- (20) Ozaki, N.; Tezuka, H.; Tezuka, H.; Inaba, M. A simple acceleration method for the Louvain algorithm. *Int. J. Electr. Comput. Eng.* **2016**, *8*, 207.
- (21) Kaufmann, K.; Vecchio, K. S. Searching for high entropy alloys: A machine learning approach. *Acta Mater.* **2020**, *198*, 178–222.
- (22) Zhang, Y.; Wen, C.; Wang, C.; Antonov, S.; Xue, D.; Bai, Y.; Su, Y. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Mater.* **2020**, *185*, 528–539.
- (23) Huang, F.; Li, X.; Zhang, S.; Zhang, J.; Chen, J.; Zhai, Z. Overlapping community detection for multimedia social networks. *IEEE Trans. Multimed.* **2017**, *19*, 1881–1893.
- (24) Eberhart, R.; Kennedy, J. A new optimizer using particle swarm theory. *MHS'95 Proceedings of the Sixth International Symposium on Micro Machine and Human Science*; IEEE, 1995; pp 39–43.
- (25) Ahajjam, S.; El Haddad, M.; Badir, H. A new scalable leader-community detection approach for community detection in social networks. *Soc. Network* **2018**, *54*, 41–49.
- (26) M'Barek, M. B.; Borgi, A.; Bedhiafi, W.; Hmida, S. B. Genetic algorithm for community detection in biological networks. *Procedia Comput. Sci.* **2018**, *126*, 195–204.
- (27) Machaka, R. Machine learning-based prediction of phases in high-entropy alloys. *Comput. Mater. Sci.* **2021**, *188*, 110244.
- (28) Zhao, X.; Liang, J.; Wang, J. A community detection algorithm based on graph compression for large-scale social networks. *Inf. Sci.* **2021**, *551*, 358–372.
- (29) Khan, B. S.; Niazi, M. A. Network community detection: A review and visual survey. **2017**, arXiv:1708.00977, arXiv preprint.
- (30) Rostami, M.; Berahmand, K.; Forouzandeh, S. A novel community detection based genetic algorithm for feature selection. *J. Big Data* **2021**, *8*, 2.
- (31) Nasiri, E.; Berahmand, K.; Rostami, M.; Dabiri, M. A novel link prediction algorithm for protein-protein interaction networks by attributed graph embedding. *Comput. Biol. Med.* **2021**, *137*, 104772.
- (32) Risal, S.; Zhu, W.; Guillen, P.; Sun, L. Improving phase prediction accuracy for high entropy alloys with machine learning. *Comput. Mater. Sci.* **2021**, *192*, 110389.
- (33) Jalali, M.; Tsotslas, M.; Wöll, C. MOFSocialNet: Exploiting Metal-Organic Framework Relationships via Social Network Analysis. *Nanomaterials* **2022**, *12*, 704.

## □ Recommended by ACS

### Improving DNA-Binding Protein Prediction Using Three-Part Sequence-Order Feature Extraction and a Deep Neural Network Algorithm

Jun Hu, Gui-Jun Zhang, *et al.*

JANUARY 31, 2023

JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ

### I21-Norm-Based Robust Feature Extraction Method for Fault Detection

Xin Sha and Naizhe Diao

NOVEMBER 25, 2022

ACS OMEGA

READ

### Sequence-Based Prediction of Plant Allergenic Proteins: Machine Learning Classification Approach

Miroslava Nedalkova, Vasil Simeonov, *et al.*

JANUARY 20, 2023

ACS OMEGA

READ

### Janus γ-GeSSe Monolayer as a High-Performance Material for Photocatalysis and Thermoelectricity

Vuong Van Thanh, Nguyen Tuan Hung, *et al.*

JANUARY 04, 2023

ACS APPLIED ENERGY MATERIALS

READ

Get More Suggestions >