

Graphical representation and hierarchical decomposition mechanism for vertex-cover solution space

Wei Wei^{a,b,c,d}, Xiangnan Feng^{e,*}

^a School of Mathematical Sciences, Beihang University, Beijing, 100191, People's Republic of China

^b Key Laboratory of Mathematics, Informatics and Behavioural Semantics, Ministry of Education, 100191, People's Republic of China

^c Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing, 100191, People's Republic of China

^d Peng Cheng Laboratory, Shenzhen, Guangdong, 518055, People's Republic of China

^e Center for Humans and Machines, Max Planck Institute for Human Development, Berlin, 14195, Germany

ARTICLE INFO

Keywords:

KE-layer structure
Minimum vertex cover
Random graphs
Solution space expression

ABSTRACT

NP problems act essential roles in modelling and analysing various complex systems, and representation learning of system individuals and relations has faced the kernel difficulty in understanding the complexity and solving the NP problems. In this paper, solution space organisation of minimum vertex-cover problem is deeply investigated using the famous König-Egervary (KE) graph and theorem, in which a hierarchical decomposition mechanism named KE-layer structure of general graphs is proposed to reveal the complexity of vertex-cover. To achieve the graphical representation and hierarchical decomposition, an algorithm to verify the KE graph is given by the solution space expression of vertex-cover, and the relation between multi-layer KE graphs and maximal matching is illustrated and proved. Furthermore, a framework to calculate the KE-layer number and approximate the minimal vertex-cover is provided, with different strategies of switching nodes and counting energy. The phase transition phenomenon between different KE-layers is studied with the transition points located, and searching of vertex-cover got by this strategy presents comparable advantage against several other methods. Its efficiency outperforms the existing ones just before the transition point. The graphical representation and hierarchical decomposition provide a new perspective to illustrate the structural organisations of graphs better, and its formation mechanism can help reveal the intrinsic complexity and establish heuristic strategy for large-scale graphs/systems recognition.

1. Introduction

As the development of the computer science and artificial intelligence, the concern on data analysis and objective optimization becomes increasingly important. Many cutting-edge technologies in machine learning and deep learning greatly promote the current research in a variety of fields [1,2]. Core difficulty faced by these technologies is commonly and mainly the high computational complexity induced by the real-world applications. Different approximating strategies on the computing-hard problems (i.e., NP even NP-Complete problems) have achieved a great success in game theory, natural language processing, image processing etc., assisted with the computer software and hardware technologies. For example, exhaustive searching in game theory can be substituted by

* Corresponding author.

E-mail address: fengxiangnan@gmail.com (X. Feng).

<https://doi.org/10.1016/j.amc.2023.128264>

Received 1 November 2021; Received in revised form 21 July 2023; Accepted 25 July 2023

Available online 2 August 2023

0096-3003/© 2023 Elsevier Inc. All rights reserved.

strategic computation using deep reinforcement learning [3], feature extraction and pattern recognition in image processing can be automatically accomplished by convolution and pooling in convolution neural network [4], and graph partition in combinatorial optimization can be calculated by node coding using graph representation learning [5]. A large range of NP and NP-Complete problems in various applications provides scenarios for the nowadays artificial intelligence technologies.

Combined with tools from statistics and graph theory, a large number of researches from graph-based complex systems (networks) [6] have been applied in many engineering and natural science fields, like biology [7], social structure [8], world wide web (WWW) [9] and human dynamics [10]. Transforming irregular complex systems into regularized organizations/representations is always an active topic in system learning, e.g., the images are treated as tensor in convolution neural network [11], and the graph nodes can be expressed as vectors in graph representation learning [12–14]. For NP or NP-Complete problems such as combinatorial optimization problems based on graphs, expressing underlying roles or relations of the nodes is also a kernel idea to solve and understand their complexity. In statistical mechanics, the nodes on a graph of combinatorial optimization problems are assigned marginal probabilities in the solution space by the replica and cavity method [15,17], which can also be viewed as nodes' representation to reveal the underlying values and relations. The representation mechanism of network/graph aims at decoupling the complicated relations among the nodes to simplify the whole system, which is generally focused on in the nowadays representation learning research [18,19]. In classical linear algebraic systems, its solution spaces can be represented by a linear space composed of some basis vectors [20], and in a class of pure nonlinear algebraic system there is still a group of basis representation for all of its solutions [21]. Meanwhile, hierarchical decomposition of networks/graph provides effective understanding of the complicated system structure and functions: centrality measurements and community detecting methods like k -shell [22], k -core [23] and modularity [24], and many leaf-removing strategies [25,15–17,26] are commonly used, which could illustrate the layer structure of graphs and point out the hubs.

Existing representation studies on the NP problems mainly concern on the node status representation. We have proposed a strict graphical representation of the whole solution space of minimum vertex-cover problem when the graph topology is bipartite [27], which combined the node statuses with the edge statuses. In this paper, the corresponding results for minimum vertex-cover problem will be strictly generalized on the famous König-Egérvary (KE) graphs. Combining with the minimum coverage requirements, a hierarchical decomposition mechanism based on the solution space expression of the KE-subgraphs will be deeply investigated to reveal the intrinsic computational hardness of the NP-Complete problem. The further research on graphical representation and hierarchical decomposition mechanism for vertex-cover solution space gives a representative instance for understanding the complexity of NP-Complete problems, and could provide insightful recognition and efficient way in analysing graph-based complex systems.

2. Preliminary of KE graph and solution space expression for vertex-cover

In this section, some basic knowledge about the König-Egérvary theorem with KE graphs and solution space expression technique for vertex-cover will be introduced as preliminary.

2.1. Relations of KE graph with minimum vertex-cover

For a graph G composed by the vertex set V with n vertexes and edge set E with m edges, a vertex-cover is a vertex subset $C(G) \subset V$ that for any edge $(v_i, v_j) \in E$ there exists either $v_i \in C(G)$ or $v_j \in C(G)$. If the cardinality of vertex-cover set achieves the lowest, it is called the *minimum vertex-cover* with the corresponding cardinality named as *minimum vertex-cover number* of a graph. For a graph there could exist more than one minimum vertex-covers [28]. The problem of finding the minimum vertex-cover belongs to Karp's 21 NP-Complete problems [29] and the 6 basic NP-Complete problems [30], which have broad applications in real world but have been recognized to be intrinsically hard to solve [31].

Another widely studied topic in graph theory is the maximum edge matching [28]. An edge matching, or independent edge set, of a graph is an edge subset $M \subset E$ satisfying that any two edges in M have no common vertex, and a vertex is matched if it is an endpoint of one edge in M . The edge matching set with the largest cardinality is called the *maximum (edge) matching* with the corresponding cardinality named as *maximum matching number* of a graph, and there could exist more than one maximum matchings for a graph. Finding the maximum matching for a general graph belongs to P problems, namely for a graph its maximum matching could be found in polynomial time. The blossom algorithm is widely applied to find the maximum matching for general graphs [32], while the Hungarian maximum matching algorithm is widely used to find the maximum matching for bipartite graphs [33]. The maximum matching has a large number of applications, e.g., this concept is deeply related to Kekulization and the process of Tautomerization in chemistry [34].

Bipartite graph can associate the minimum vertex-cover with its maximum matching. A graph G is *bipartite* if there exist two non-empty node subsets V_1 and V_2 with $V_1 \cap V_2 = \emptyset$ and $V_1 \cup V_2 = V$, satisfying that for any edge $(v_i, v_j) \in E$, v_i and v_j must belong to different subsets V_1 or V_2 . The bipartite graph could be used to illustrate the relation between objects with two different types especially in social network study [35], e.g., the relations between actors and movies. One of the most crucial property of bipartite graph, given by König-Egérvary theorem, is that its maximum matching number is equal to its minimum vertex-cover number. This property builds a bridge between the maximum matching problem and minimum vertex-cover problem and makes it possible to find the minimum vertex-cover of bipartite graphs in polynomial time [36].

A graph whose maximum matching number equals to its minimum vertex cover number is called the König-Egérvary (KE) graph. There exist more types of KE graphs than the bipartite ones, and a commonly studied one is the No-Leaf-Removal-Core graph. In minimum vertex-cover problem, a leaf is regarded as a one-degree node and its neighbor with their connecting edge [16], which

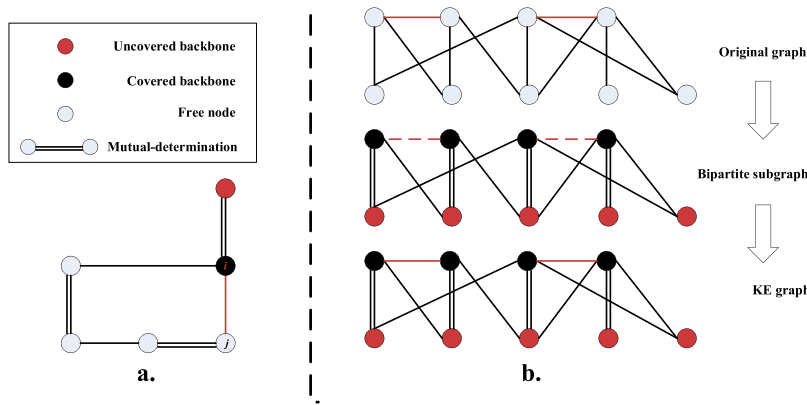


Fig. 1. A schematic view of the graphical representation and hierarchical decomposition mechanism for vertex-cover solution space. Only double edges between free nodes on the reduced solution graph represent mutual-determinations, and others correspond to a group of maximum matching in the absence of ambiguity. a shows an example of exact solution space expression; b illustrates how to achieve a solution space expression for a KE graph.

suggests exactly one cover number and at most one edge matching in a leaf. If there exist only isolated nodes or no node after all the leaves are removed iteratively, this graph is called *No-Leaf-Removal-Core graph* [37]. For the No-Leaf-Removal-Core graph, its minimum vertex-cover number equals to its leaf number, and since different leaves are independent subgraphs (no common nodes and edges), its maximum matching number should not be smaller than its leaf number; for the vertex cover of a graph, each independent edge should occupy at least one cover number, and the minimum vertex-cover number must not be smaller than its maximum matching number. Thus, the No-Leaf-Removal-Core graph should have its maximum matching number equal to its minimum vertex-cover number, and it is one type of KE graph.

2.2. Solution space expression of vertex-cover

For the minimum vertex-cover problem, our aim is to locate the least number of covered nodes and make each edge on the given graph touched by at least one such node. Generally, one given graph owns many minimum vertex-covers which form a solution space for this combinatorial optimization problem.

From the former work [38], it has been already known that on structures of KE graphs including trees, no leaf-removal-core graphs and bipartite graphs, for minimum vertex-cover problem, we have corresponding algorithms to achieve the solution space expression, which is called *reduced solution graph*. The reduced solution graph provides a complete description of all the minimum vertex-cover solutions, and it can be easily obtained on these topologies [27]. On the reduced solution graph expression, all the nodes are classified as *uncovered backbones*, *covered backbones* and *free nodes*; nodes should always be uncovered or covered separately in all the solutions, namely *frozen* nodes, or could have alternative assignments on different solutions, namely *unfrozen* nodes; all edges are classified as single or double edges, in which double edges suggest the *mutual-determination* relations between two unfrozen nodes and these two nodes will mutually affect the values of each other (two free nodes in a mutual-determination relation must have one and only one covered node) and the single edges still follow the coverage requirement.

Using the representation of the three classes of nodes and the two types of edges on the original graph, the vertex-cover solution space could be sufficiently expressed for a large class of graph instances, and indeed most of the known easily-solving minimum vertex-cover instances belong to KE graphs [38,27], which have exact solution space expressed by the node and edge class labels. In Fig. 1, a schematic view of the solution space expression is provided. Fig. 1a shows the different classes of nodes and edges, and gives an instance for minimum vertex-cover solution space of a simple graph. In this graph, node *i*, marked by black, has to be covered in all solutions, which makes its one-degree neighbor node, marked by red, namely uncovered all the time; the double edges connect the free nodes, and since on each double edge, one node getting covered will make the other one uncovered, the two double edges in the graph lead to two covered nodes. Thus in this case, the minimum vertex-cover number of the graph is 3, and all possible solutions are presented by the marked nodes and different type of edges. Fig. 1b is an example of getting a solution space expression for a KE graph, of which detailed procedures are introduced in [27].

3. Solution space expression algorithm for vertex-cover of KE graph

In this section, an algorithm based on the solution space expression will be provided to judge whether a graph is a KE graph. Meanwhile, the vertex-cover solution space expression for KE graph can be obtained accordingly.

Since the edges in the maximum matching are independent and all the covered nodes must be matched nodes in a KE graph, there must be one and only one covered node in each matching edge, and all the unmatched nodes must be uncovered backbones in the minimum vertex-cover solution space of KE graph. Thus, there is a natural correspondence between the concept of edge matching and the mutual-determination relation on KE graphs, which also could be used to determine whether a graph is a KE one. The algorithm to examine the KE graph and get the solution space expression of its minimum vertex-cover is illustrated in Algorithm 1, which involves the following steps:

Algorithm 1 Get Solution Space Expression for KE Graph Vertex-Cover.

Input: Graph G ; **Output:** Reduced solution graph $S(G)$ of G

```

M = Maximum matching of G, Initialize  $S(G) = G$ ;
for Any match  $(i,j)$  in M do
    Set  $(i,j)$  as a double edge in  $S(G)$ ;
end for
while Any unmatched node  $k$  exists do
    Set  $k$  as uncovered backbone in  $S(G)$ ;
     $S(G) = \text{Freezing-Influence}(S(G), k)$ ;
end while
end while
 $S(G) = \text{Conflict-Checking}(S(G))$ ;
*  $S(G) = \emptyset$  means graph  $G$  is not a KE graph.
    
```

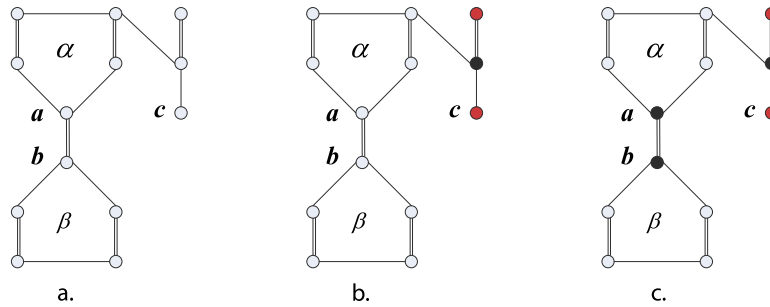


Fig. 2. An instance for determining whether a graph is a KE one, with red nodes for uncovered backbones, black nodes for covered backbones and white nodes for free nodes. **a** is the original graph with its maximum matching from **Step 1**. **b** gives an unmatched node c and performs a *Freezing-Influence* operation from **Step 2**. **c** checks the nodes in the two cycles α, β , and finds that nodes a and b must be both covered backbones by **Step 3**. But nodes a and b form a matching with two covered nodes, which conflicts with the KE property, and in this instance it is not a KE graph.

Step 1: For the graph, find one of its maximum matching, which could be done in polynomial time. If the graph is a KE one, then all the covered nodes should be matched.

Step 2: Assign double edges for each matched edge and mark the unmatched nodes as uncovered backbones, do the *Freezing-Influence* described in Algorithm 2 on the graph, which makes sure that all the neighbors of an uncovered backbone should be covered backbones, and the mutual-determination neighbors (double edges) of a covered backbone should be an uncovered backbone. If there is some *conflict* while determining the state of some nodes (i.e., some nodes are required to be covered by one neighbor and uncovered by another simultaneously), the cover number of the whole graph will exceed the maximum matching number and it cannot be a KE graph.

Step 3: Apply the *Conflict-Checking* from Algorithm 3 for the rest unfrozen nodes after **Step 2**, which involves the *Freezing-Influence* and *Consistency-Checking* in Algorithm 2. Considering unfrozen node i , set i as covered and uncovered separately and check whether there is conflict: if there are conflicts on both cases, then the graph is not a KE graph and additional cover number is inevitable; if there is only conflict on one case (say the covered case), then node i should be kept away from this case (it should be an uncovered backbone); if there is no conflict on both cases, then node i should be a free node. For all the new produced backbones, the *Freezing-Influence* operation should be also performed.

Step 4: If the process of **Step 3** can survive after all the nodes are checked, it is a KE graph and its solution space expression of minimum vertex-cover could be got. Otherwise it is not a KE graph.

For the above process, conflict analysis is the key to determine whether the graph is a KE one. All the conflicts could come from **Step 2** and **Step 3**. For the conflict in **Step 2**, it results from the KE property and requirement of minimum coverage, and it is an exact step. For the conflict in **Step 3**, node i cannot have values ensuring the KE property for both conflict cases; for the single conflict case, it indeed fulfills the *odd cycle breaking* [38], which makes it also an exact step. An instance is shown in Fig. 2.

For the obtained reduced solution graph after **Steps 1-4**, an important question is whether any undiscovered conflict exists when different free nodes are assigned values? By the solution space expression of minimum vertex-cover, if one free node i is applied the checking processes in **Steps 1-4**, it could be easily proved that node i must be a free node in the whole solution space by the following process: set node i as uncovered backbone and do the *Freezing-Influence* propagation; if it passes **Step 3**, it can fix the states of some nearby nodes and no conflict exists, and then the solution space shrinks; choose any residual free node to fix its state and repeat the operation recursively, a complete minimum coverage solution can be obtained; similarly, set i as covered backbone and another complete solution can also be obtained. Therefore, node i is proved to have a free state in the solution space.

The above process provides a way to achieve the whole vertex-cover solution space of a KE graph, and it ensures that one reduced solution graph can represent all the possible solutions. This process costs time only on the maximum matching and *Conflict-Checking*. Except the locating of maximum matching, the time complexity will cost no more than $O(n)$.

Algorithm 2 KE Graph Verification Functions.

```

function Freezing-Influence( $S(G), k$ )
  for Any  $k$ 's neighbor  $j$  do
    Set  $j$  as covered backbones in  $S(G)$ ;
  end for
  while Any double edge  $(i, j)$  has node  $i$  covered backbone and  $j$  unfrozen do
    Make  $j$  uncovered backbone in  $S(G)$ ;
     $S(G) = \text{Freezing-Influence}(S(G), j)$ ;
  end while
  Return  $S(G)$ 
end function

function Consistency-Checking( $S(G)$ )
  for Any uncovered backbone  $i$  in  $S(G)$  do
    if There exist uncovered backbone neighbors of  $i$  then
      Return 1
    end if
  end for
  for Any double edge in  $S(G)$  do
    if Its two ends are both covered backbones then
      Return 1
    end if
  end for
  Return 0
end function

```

Algorithm 3 Conflict Checking Process.

```

function Conflict-Checking( $S(G)$ )
  Consistency-Checking( $S(G)$ )
  for Any unfrozen node  $i$  in  $S(G)$  do
    Set  $S^{++}(G) = S(G)$ , and set  $i$  as uncovered backbone in  $S^{++}(G)$ ;
     $S^{++}(G) = \text{Freezing-Influence}(S^{++}(G), i)$ ;
     $posi = \text{Consistency-checking}(S^{++}(G))$ ;
    Set  $S^{--}(G) = S(G)$ , and set  $i$  as covered backbone in  $S^{--}(G)$ ;
    if  $i$  has a double-edge free neighbor  $j$  then
      Set  $j$  as uncovered backbone in  $S^{--}(G)$ ;
       $S^{--}(G) = \text{Freezing-Influence}(S^{--}(G), j)$ ;
       $nega = \text{Consistency-Checking}(S^{--}(G))$ ;
    end if
    if  $posi = 1$  and  $nega = 1$  then
       $S(G) = \emptyset$ , break;
    end if
    if  $posi = 1$  and  $nega = 0$  then
       $S(G) = S^{--}(G)$ ;
    end if
    if  $posi = 0$  and  $nega = 1$  then
       $S(G) = S^{++}(G)$ ;
    end if
    Return  $S(G)$ 
  end for
end function

```

4. Multi-layer KE subgraph decomposition for vertex-cover

4.1. Definition of multi-layer KE graph

By the definition of the KE graph whose minimum vertex-cover number is equal to its maximum matching number, there are strong relations between the KE graph and its maximum matching. Here, we classify all the nodes into two separated classes A_1 and B_1 with $A_1 \cup B_1 = V$ and $A_1 \cap B_1 = \emptyset$ based on the maximum matching, where it is required that every two nodes on a same matching edge cannot be assigned into the same class and the unmatched nodes are all placed in class B_1 without loss of generality. Evidently there are exponential ways of arranging nodes into A_1 and B_1 . The following result provides an alternative understanding of KE graphs.

Theorem. For any KE graph G , there must exist some arrangements of A_1^* and B_1^* in which B_1^* has no inner-class edges, i.e., all the edges in G could only be inner-class edges in A_1^* or the inter-class edges between A_1^* and B_1^* , and any two matched nodes on the same matching edge could not belong to the same class. Correspondingly, a graph satisfying some arrangements of A_1^* and B_1^* above by maximum matching in which B_1^* has no inner-class edges, must be a KE graph.

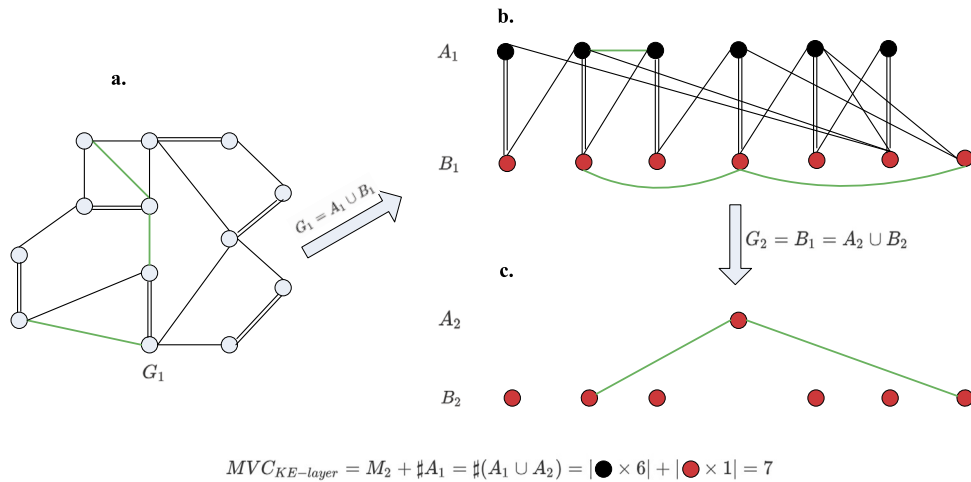


Fig. 3. A 2-layer KE graph example and its decomposing process. a is the original graph with one maximum matching marked by double-edge. b gives a classification in which nodes are assigned into A_1 and B_1 ; nodes on the same matched edge are not assigned into the same set and all unmatched nodes are placed into B_1 (only one node in this case). There are two inner-class edges in B_1 . In c the G_2 is the subgraph induced by nodes in B_1 and it could be decomposed into A_2 and B_2 . There is no edge in B_2 in this case, which suggests that subgraph induced by B_2 is a KE graph and the vertex-cover of the original graph is consisted of A_1 and A_2 . The cover number of the graph in a got by this strategy is 7, which exactly equals to the minimum cover number of this graph.

Proof. \implies For a KE graph G , since its minimum vertex-cover number is equal to its maximum matching number, all the covered nodes must be the matched nodes and each matching edge has one and only one covered nodes. Choosing all the covered nodes in a minimum vertex-cover to form class A_1^* and the rest assigned into B_1^* , all the nodes in A_1^* should make the coverage of all the edges in G , and it suggests that there is no inner-class edge in the class B_1^* . Furthermore, if there are different solutions of the minimum vertex-cover of graph G , there will be different arrangements of A_1^* and B_1^* .

\Leftarrow If a graph G has the arrangement of A_1^* and B_1^* where B_1^* has no inner-class edges, all nodes in A_1^* will lead to a vertex-cover of G . As the nodes number of A_1^* is equal to the maximum matching number, it also determines the lower bound of the minimum coverage. Thus, the nodes in A_1^* form a minimum vertex-cover, which results in G being a KE graph. \square

By the above theorem, a graph is a KE graph if and only if after deleting all the nodes in A_1^* with related edges, the rest graph related with nodes in B_1^* is a graph consisting of isolated nodes without any edge. In this paper, we also call a classical KE graph as 1-layer KE-graph. If a graph G is not a 1-layer KE-graph, it implies that there exists no such B_1^* that does not contain inner-class edges. In this case, our aim will be adjusted to find the proper separated classes A_1^* and B_1^* , in which the subgraph induced by B_1^* has the lowest coverage number (smallest number of covered nodes) with respect to the minimum vertex-cover. Evidently, the KE graph can be viewed as a special case with the subgraph induced by B_1^* having 0 coverage in this setting.

However, the core difficulty is to achieve the decomposition of A_1^* and B_1^* , i.e., to find the subgraph induced by B_1^* having the lowest coverage. In the proper classification A_1^* and B_1^* , the subgraph induced by A_1^* with the inter-class edges should occupy the coverage of vertex-cover the same as $\#A_1^*$ (nodes number in A_1^*) to make as many covered nodes of G as possible belong to A_1^* ; if some node in A_1^* does not need to be covered, its matched node in B_1^* can switch with it, namely switching their class labels. As a result, the decomposition of A_1^* and B_1^* is equivalent to obtain a subgraph with coverage number $\#C(G) - \#M(G)$, where $\#C(G)$ is the minimum coverage of G and $\#M(G)$ is the maximum matching number of G . Since that the minimum vertex-cover problem belongs to NP-Complete problems and maximum matching belongs to P problems, the proper classification of A_1^* and B_1^* for general graphs is also a hard problem. Fortunately, it is easy and straightforward when the graph is a KE one, and the algorithm will be provided in next section.

Then, based on the above definition and analysis, if the subgraph G_1 induced by B_1^* is not empty and also a KE one, the graph G is called a 2-layer KE-graph, and this subgraph could find a proper separated classes A_2^* and B_2^* , in which $B_1^* = A_2^* \cup B_2^*$ and $A_2^* \cap B_2^* = \emptyset$, A_2^* and B_2^* have the same meaning with A_1^* and B_1^* but restricted on G_1 . When a graph is a 2-layer KE-graph, its minimum coverage is equal to the number of nodes in $A_1^* \cup A_2^*$. Similarly, we can define a k -layer KE-graph and nodes in G can be decomposed as $A_1^* \cup A_2^* \dots \cup A_k^* \cup B_k^*$ with the class B_k^* having no inner-class edges. In this way, the minimum coverage should be equal to $\#A_1^* \cup A_2^* \dots \cup A_k^*$. The upper bound of the k -layer KE-graph satisfies $k \leq \log_2 n$, where n is the node number of the whole graph and the upper bound will be achieved when the graph is a complete graph. An example of 2-layer KE graph with detailed decomposing process to get its minimum vertex-cover is given in Fig. 3.

4.2. Algorithmic framework for measuring KE-layer

In this section, a framework to find the KE-layer number for general graphs is proposed. Several approximation strategies will be presented to approach the KE-layer number and minimum vertex-cover of graphs. To measure the KE-layer number of a general

graph, the maximum matching needs to be calculated at first and nodes are assigned into A_1 and B_1 as required above. Then the matched pairs of nodes will be switched to lower the coverage of B_1 to approach B_1^* . The framework is described in Algorithm 4.

Algorithm 4 Framework for Measuring KE-layer.

```

Let  $l = 0, A_0 = \emptyset, B_0 = G$ ;
while  $B_l$  is not KE graph do
     $l = l + 1, G_l =$  subgraph induced by  $B_{l-1}$ ;
    Calculate the maximum matching  $M(G_l)$  of graph  $G_l$ ;
    Place each pair of matched nodes into  $A_l$  and  $B_l$  separately;
    Place every unmatched node into  $B_l$ ;
     $E = n, (E', A_l, B_l) = \text{Switch-Matched-Nodes}(G_l, A_l, B_l)$ ;
    while  $E > E'$  do
        Let  $E = E', (E', A_l, B_l) = \text{Switch-Matched-Nodes}(G_l, A_l, B_l)$ ;
    end while
end while
 $l = l + 1, G_l =$  subgraph induced by  $B_{l-1}$ ;
Calculate the maximum matching  $M_l$  of the subgraph induced by  $B_l$ ;
Return KE-layer number  $l$ , vertex-cover number  $\#M_l + \sum_{i=1}^{l-1} \#A_i$ 
**  $\text{Switch-Matched-Nodes}(G_l, A_l, B_l)$  returns new  $A_l, B_l$  and coverage of set  $B_l$ .
    
```

In this framework, for a graph with KE-layer number L , A_l with $1 \leq l \leq L$ is peeled layer by layer until the remained subgraph G_L induced by nodes in B_{L-1} is a KE one. All nodes contained in A_l and the minimum vertex-cover of G_L will compose the vertex-cover of graph G . Since G_L is a KE graph, it is fast to find its minimum vertex-cover. Thus, the vertex-cover number of graph G with KE-layer L could be calculated as:

$$MVC_{KE-layer} = \#M_L + \sum_{i=1}^{L-1} \#A_i. \tag{1}$$

During the whole process, the judgment of KE graphs on each G_l could be implemented by Algorithm 1 proposed above. An detailed process to get its minimum vertex-cover could also be seen in Fig. 3.

4.3. Strategies for switching matched nodes

A key problem in this framework is how to place the matched pairs of nodes into A_l and B_l to lower the coverage of B_l for each graph G_l , namely the *Switch-Matched-Nodes* function above. Even with the constraint that every pair of matched nodes are placed into different sets, for a graph with maximum matching number M , there are 2^M possible arrangements, which makes it hard to find the optimal solution. This problem could be described as:

$$\min_{n \in A_l \text{ or } B_l, \forall n \in V(G_l)} \text{Coverage}(G_l), \tag{2}$$

$$s.t. \forall \text{ matched edge } (i, j), i \in A_l, j \in B_l \text{ or } j \in A_l, i \in B_l. \tag{3}$$

To solve it, several methods are proposed to optimize the solution. Two factors will determine the efficiency of this framework: the switching strategy of pairs of matched nodes and how to calculate the coverage of B_l .

4.3.1. Switching strategy

A directed idea is to switch pairs of matched nodes in a greedy way. For each matched nodes pair, if switching them will lower the coverage of G_l , this switching should be kept; if the switching does not change or increase the coverage, this pair of nodes should be switched back. The detailed procedure is presented in Algorithm 5, Strategy 1.

Algorithm 5 *Switch-Matched-Nodes*(G, A, B), Strategy 1.

```

 $E =$ Coverage of  $B$ ;
for every matched nodes pair  $(i, j)$  with  $i \in A$  and  $j \in B$  do
    Place  $i$  in  $B$  and  $j$  in  $A$ , and calculate  $E' =$  Coverage of  $B$ ;
    if  $E' \geq E$  then
        Place  $i$  back in  $A$  and  $j$  back in  $B$ ;
    else
         $E = E'$ ;
    end if
end for
Return  $E, A, B$ 
    
```

Obviously, this method is flawed: lowering the coverage in every step does not mean the global optimal solution. Another idea is to introduce some uncertainty and for each step, and here two pairs of matched nodes will be selected randomly and switched to lower the coverage. This process is presented in Algorithm 6, Strategy 2.

Algorithm 6 Switch-Matched-Nodes(G, A, B), Strategy 2.

```

temp = 0;
while temp < #M(G) do
    temp = temp + 1;
    Calculate E = Coverage of B;
    Randomly select two matched nodes pairs (i1, j1), (i2, j2), with i1, i2 ∈ A, j1, j2 ∈ B;
    Place i1, i2 in B and j1, j2 in A, calculate E' = Coverage of B;
    if E' ≥ E then
        Place i1, i2 back in A and j1, j2 back in B;
    else
        E = E';
    end if
end while
Return E, A, B

```

Ensuring that each switch lowers the coverage is not always a good strategy. Sometimes, to allow the increase of coverage at several steps may bring more possibility to the solution. So in the next strategy, when the switch does not lower the coverage, we keep the switch with a certain small probability and continue the process. This strategy is applied based on **strategy 2** and is described in Algorithm 7, **Strategy 3**.

Algorithm 7 Switch-Matched-Nodes(G, A, B), Strategy 3.

```

temp = 0;
while temp < #M(G) do
    temp = temp + 1;
    Calculate E = Coverage of B;
    Randomly select two matched nodes pairs (i1, j1), (i2, j2), with i1, i2 ∈ A, j1, j2 ∈ B;
    Place i1, i2 in B and j1, j2 in A, calculate E' = Coverage of B;
    Generate a random number R;
    if E' ≥ E and R ≤ Threshold then
        Place i1, i2 back in A and j1, j2 back in B;
    else
        E = E';
    end if
end while
Return E, A, B

```

4.3.2. Coverage estimation

The coverage here refers to the minimum vertex-cover number of G_I , which is hard to get for a general graph. To estimate the coverage, a direct idea is to consider the edge number contained in G_I . For random graphs, when the average degree increases, with a high probability the vertex-cover number will increase. Another idea is to use the maximum matching number of G_I to approximate its coverage. The maximum matching number of a graph determines its lower boundary of minimum vertex-cover number, which suggests that it is a good estimation for the coverage of the graph. When the maximum matching number is high, with a high probability, this graph owns a high minimum vertex-cover number. In the following sections, we use *edge energy* and *maximum matching energy* to denote these two coverage estimations.

5. Experiments and results

In this section, a series of experiments will be conducted on random graphs. The switching strategies and vertex-cover estimating methods mentioned above will be implemented and compared. The KE-layer number and vertex-cover number of these graphs will be discussed.

5.1. KE-layer number

The KE-layer numbers of ER random graphs with different average degrees are calculated and discussed in this subsection. Each graph contains 1,000 nodes and for each average degree value, 100 graphs are generated and calculated. The average KE-layer numbers at each average degree are presented in Fig. 4. As we could see, although different strategies perform different efficiency, there is a clear transition from 1-layer KE-graph phase to 2-layer KE-graph phase. The transition point is in the degree interval (2, 3). When average degree is higher than 12, the KE-layer number of almost all the graphs reaches 3. By the results in Fig. 4, **Strategy 1** and **2** with the edge energy have good performance and they will be analyzed in the following.

The phase transition points will be estimated based on this fact: when the graphs are far from the phase transition points, most of them should have similar KE-layer number and the fluctuation (standard deviation) of their KE-layer numbers should be small, but the fluctuation should get increased when it approaches the transition points and reaches the highest at the transition point. For the location of the 1-layer KE-graph to 2-layer KE-graph transition point, by observing the standard deviation values in Fig. 5 from **Strategy 1** and **Strategy 2** with edge energy, it could be found that the transition point is around e (average degree), where

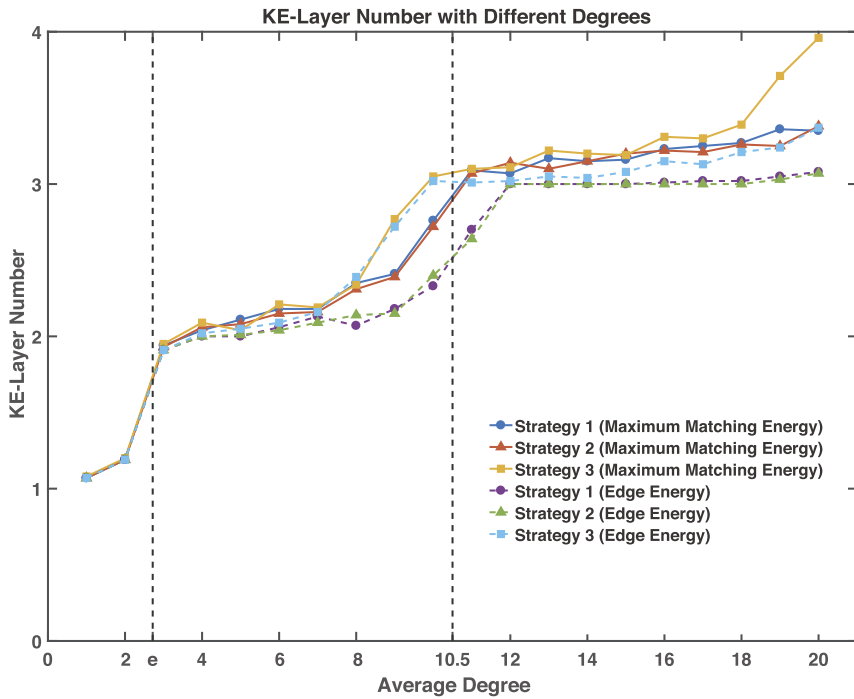


Fig. 4. The KE-layer numbers by the proposed strategies on ER graphs with average degrees ranging from 1 to 20. For each average degree value, 100 graphs with 1000 nodes are generated. Results of Strategy 1, 2, 3 with both edge energy and maximum matching energy are drawn on the graph. The possible position phase transition points from KE-level 1 to KE-level 2 and from KE-level 2 to KE-level 3 are drawn on dotted line.

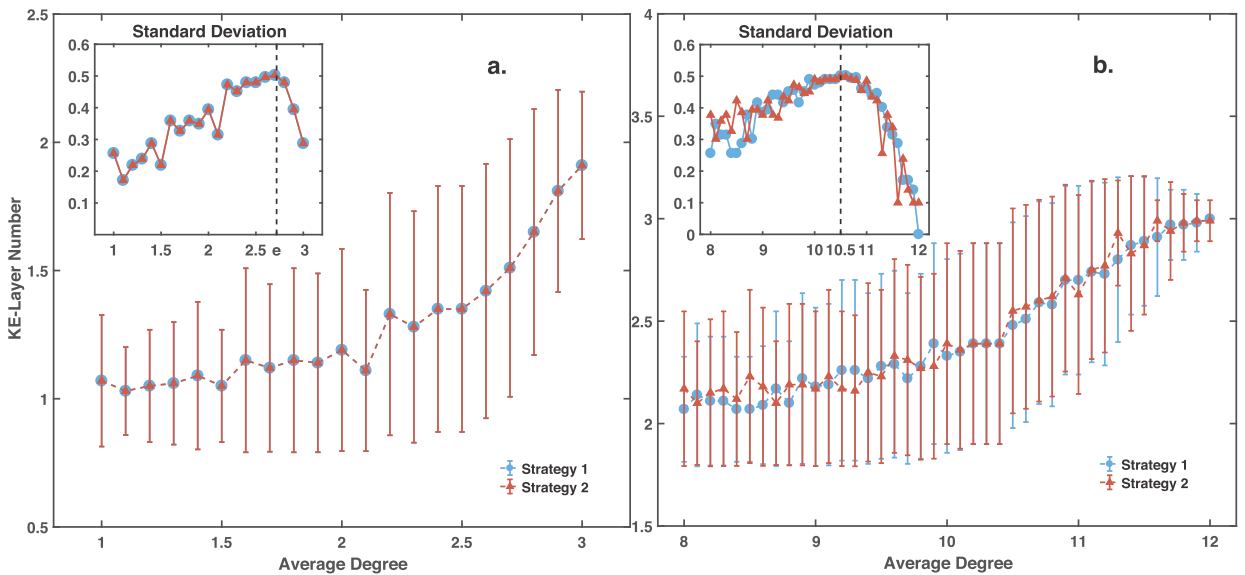


Fig. 5. The KE-layer number errorbars by the proposed methods on ER graphs with various average degrees. For each average degree value, 100 graphs with 1,000 nodes are generated. Results of Strategy 1, 2 with edge energy are drawn on the graph. The values of standard deviation are plotted in the inside figures. a. KE-layer numbers on ER graphs with average degrees ranging from 1 to 3. b. KE-layer numbers on ER graphs with average degrees ranging from 8 to 12.

the standard deviation reaches the highest and gets lower quickly after. For the location of the 2-layer KE-graph to 3-layer KE-graph transition point, a similar analysis could be conducted and the phase transition point is around 10.5 (average degree). Also when the 3-layer KE-graph starts to get dominate (average degree larger than 12), the standard deviation is very close to zero, which suggests that almost all the graphs with this average degree are 3-layer KE-graphs. Certainly, there will be more k -layer KE-graphs and corresponding phase transitions with $k > 3$ when the average degree gets increased over 20, which will not be discussed in this paper.

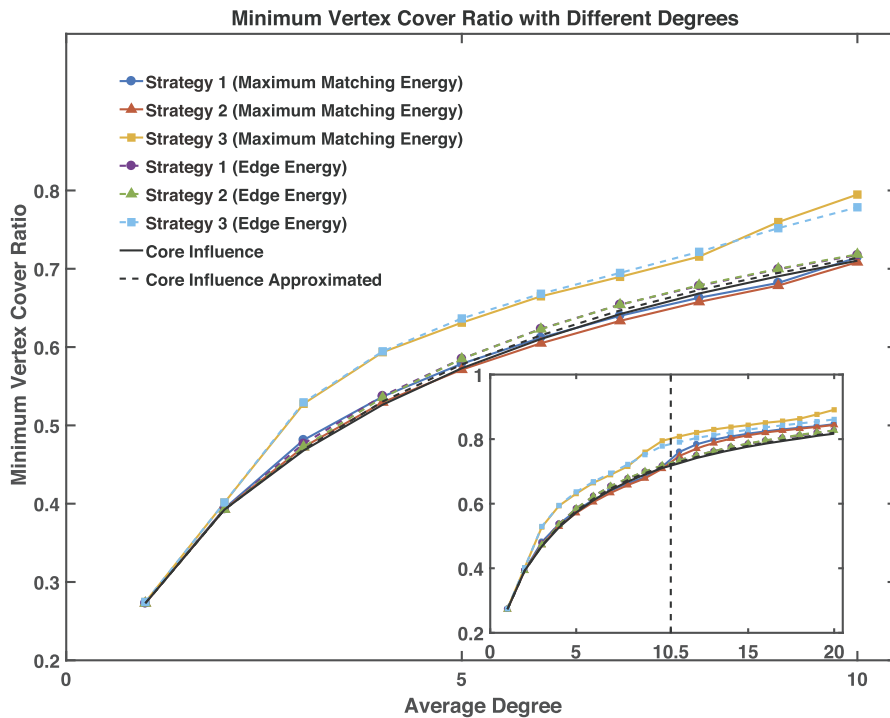


Fig. 6. The vertex-cover ratio by the methods we proposed. The experiments are conducted on ER graphs with different average degrees. For each average degree value, 100 graphs are generated and each graph contains 1,000 nodes. Results of Strategy 1, 2, 3 with both edge energy and maximum matching energy, and core influence are drawn on the graph. The vertex-cover ratio with average degrees from 0 to 20 are plotted in the inside figure. The transition point 10.5 is plotted, around which the vertex-cover ratios of Strategy 1 and 2 with edge energy surpass those with maximum matching energy.

5.2. Minimum vertex-cover number

The process of calculating the KE-layer number could also approximate the minimum vertex-cover for the graphs. The experiments are also conducted on ER graphs, and in Fig. 6 the vertex-cover numbers of different strategies mentioned above are presented. As the Figure shown, before around average degree 10, the Strategy 2 with maximum matching energy works best compared to others. When the edge number contained in G_l is used to estimate the coverage, results from Strategy 1 and Strategy 2 are quite similar, and perform the best when the average degree is higher than 10. An interesting phenomenon is that results of edge energy surpass the results of maximum matching energy at around degree 10. The reason is that when the average degree is higher than 10, Strategy 1 and 2 with edge energy could get lower KE-layer numbers, which lower the vertex-cover numbers at the same time. We also compare the results above with vertex-cover numbers from core influence [37] in Fig. 6. It could be observed that the KE-layer method performs better than core influence and could get lower minimum vertex-cover number with average degree lower than 10, and it performs almost same well with core influence when average degree is higher than 10.

To further study the efficiency of our KE-layer strategy, the vertex-cover numbers results from Strategy 2 with maximum matching energy are compared with the exact minimum vertex-cover numbers on small-scale random graphs. The differences between $MVC_{KE-layer}$ and real minimum vertex-cover numbers are calculated and the ratio of these differences against the total node numbers are presented in Table 1. As we could see, the KE-layer strategy performs well and when numbers of nodes increase the cover numbers stay stable. When the edge numbers increase with fixed node numbers, the gaps become larger, which is because of the increase of densely-connected clusters. Yet the KE-layer strategy still performs high efficiency and the vertex-cover numbers stay close to the real minimum vertex-cover numbers.

6. Conclusion and discussion

In this paper, the graphical representation and hierarchical decomposition mechanism of Vertex-Cover are studied and based on this, the KE-layer structure for general graphs is proposed. A sufficient and necessary condition of determining KE graphs is given in the viewpoint of classifying the nodes into different classes based on the maximum matching, and an algorithm for verifying the KE graphs is provided, which reveals the structural features and allows us to explore the hierarchical layer structure and the complexity of vertex-cover solution space evolution. Framework for calculating the KE-layer number for general graphs is discussed, and to solve the arranging and switching problem, several algorithmic strategies are compared including the switching strategies and coverage estimating methods. The phase transition points from 1-layer KE-graph to 2-layer KE-graph and from 2-layer KE-graph to 3-layer KE-graph are also located. At the same time, in this process, the vertex-cover number could be calculated and approximated,

Table 1

The percentages of differences between results of vertex-cover number from **Strategy 2** with maximum matching energy and exact minimum vertex-cover numbers against node numbers. The experiments are conducted on the ER graphs with 80, 100 and 120 nodes with average degrees from 3 to 7 and every result is the average of 30 graphs.

Node Numbers	Average Degrees				
	3	4	5	6	7
$N = 80$	0.88%	1.38%	1.83%	1.42%	1.92%
$N = 100$	0.97%	1.77%	1.60%	1.63%	1.73%
$N = 120$	0.47%	1.44%	1.61%	1.81%	1.58%

experiments are conducted on ER graphs to examine the efficiency, and the performance is related to the KE-layer complexity. This research provides a new perspective to approach the complexity of graphs and minimum vertex-cover problem.

A lot of research on this topic can be expected in future research. The switching methods could be further explored to improve the efficiency. Also, calculating the maximum matching for each step is very time-consuming, new convenient and fast coverage estimation measurement indexes deserve more exploration. At the same time, more other aspects related to this research could be expected. For example, some microscope structure, like motifs [39] and graphlets [40], may play crucial roles in the formation of KE graphs. Finding these structural components could improve the efficiency greatly. Also, relations between minimum vertex-cover and KE graphs could be further studied to find more frameworks to approach the complexity of NP problems.

CRedit authorship contribution statement

Wei Wei: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Xiangnan Feng:** Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported in part by the Research and Development Program of China (No. 2018AAA0101100), the National Natural Science Foundation of China (No. 62050132), and the Beijing Natural Science Foundation (China) (No. 1192012, Z180005).

References

- [1] J. Friedman, T. Hastie, R. Tibshirani, in: *The Elements of Statistical Learning*, in: Springer Series in Statistics, vol. 1(10), Springer, New York, 2001.
- [2] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016.
- [3] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, *Mastering the game of go without human knowledge*, *Nature* 550 (7676) (2017) 354–359.
- [4] S. Lo, H.P. Chan, J.S. Lin, H. Li, M.T. Freedman, S.K. Mun, *Artificial convolution neural network for medical image pattern recognition*, *Neural Netw.* 8 (7–8) (1995) 1201–1214.
- [5] R. Sato, M. Yamada, H. Kashima, *Approximation ratios of graph neural networks for combinatorial problems*, 2019.
- [6] M.E.J. Newman, *Networks: An Introduction*, Oxford Univ. Press, 2000.
- [7] M. Zitnik, M. Agrawal, J. Leskovec, *Modeling polypharmacy side effects with graph convolutional networks*, arXiv:1802.00543, 2018.
- [8] D. Centola, J. Becker, D. Brackbill, A. Baronchelli, *Experimental evidence for tipping points in social convention*, *Science* 360 (6393) (2018) 1116–1119.
- [9] C. Orsini, E. Gregori, L. Lenzi, D. Krioukov, *Evolution of the Internet k -dense structure*, *IEEE/ACM Trans. Netw.* 22 (6) (Dec. 2014) 1769–1780, <https://doi.org/10.1109/TNET.2013.2282756>.
- [10] L. Alessandretti, P. Sapiezynski, V. Sekara, S. Lehmann, A. Baronchelli, *Evidence for a conserved quantity in human mobility*, *Nat. Hum. Behav.* 2 (7) (2018) 485.
- [11] H.J. Yoo, *Deep convolution neural networks in computer vision: a review*, *IEIE Trans. Smart Proces. Comput.* 4 (1) (2015) 35–43.
- [12] T.N. Kipf, M. Welling, *Semi-supervised classification with graph convolutional networks*, 2016.
- [13] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio', Y. Bengio, *Graph attention networks*, arXiv:1710.10903, 2018.
- [14] F. Chen, Y. Wang, B. Wang, C.J. Kuo, *Graph representation learning: a survey*, *APSIPA Trans. Signal Inf. Proces.* 9 (2020).
- [15] M. Mezard, F. Ricci-Tersenghi, R. Zecchina, *Two solutions to diluted p-spin models and XORSAT problems*, *J. Stat. Phys.* 111 (3–4) (2003) 505–533.
- [16] M. Weigt, A.K. Hartmann, *Minimal vertex covers on finite-connectivity random graphs: a hard-sphere lattice-gas picture*, *Phys. Rev. E* 63 (5) (2001) 056127.
- [17] S. Mertens, M. Mézard, R. Zecchina, *Threshold values of random K-SAT from the cavity method*, *Random Struct. Algorithms* 28 (3) (2006) 340–373.
- [18] Forsyth, David, *Representation learning*, *Computer* 48 (4) (2015) 6.
- [19] W.L. Hamilton, R. Ying, J. Leskovec, *Representation Learning on Graphs: Methods and Applications*, 2017.
- [20] S. Lang, *Algebra*, Addison-Wesley, Reading, MA, 1984.
- [21] B. Guo, W. Wei, Y. Sun, Z. Zheng, *Algebraic characteristics and satisfiability threshold of random Boolean equations*, *Phys. Rev. E* 81 (2010) 031122.
- [22] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H.A. Makse, *Identification of influential spreaders in complex networks*, *Nat. Phys.* 6 (11) (2010) 888.
- [23] S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, *K-core organization of complex networks*, *Phys. Rev. Lett.* 96 (4) (2006) 040601.

- [24] M.E. Newman, Modularity and community structure in networks, *Proc. Natl. Acad. Sci.* 103 (23) (2006) 8577–8582.
- [25] B. Bollobas, *Random Graphs*, Academic Press, New York, 1985.
- [26] W. Wei, B. Guo, Z. Zheng, Statistical and algebraic analysis of a family of random Boolean equations, *J. Stat. Mech. Theory Exp.* 2009 (02) (2009) P02010.
- [27] W. Wei, Y. Zhang, T. Wang, B. Li, B. Niu, Z. Zheng, Research on solution space of bipartite graph vertex-cover by maximum matchings, *J. Stat. Mech. Theory Exp.* 2015 (11) (2015) P11027.
- [28] N. Biggs, E.K. Lloyd, R.J. Wilson, *Graph Theory*, Oxford University Press, 1986, pp. 1736–1936.
- [29] R.M. Karp, in: *Proc. Sympos. IBM Thomas J. Watson Res, Center Plenum, New York, 1972*, pp. 85–103.
- [30] S.A. Cook, in: *Proc. 3rd Ann. ACM Symp. on Theory of Computing, Association for Computing Machinery, New York, 1971*, pp. 151–158.
- [31] C.H. Papadimitriou, *Computational Complexity*, John Wiley and Sons, Chichester, 2003.
- [32] J. Edmonds, Paths, trees, and flowers, *Can. J. Math.* 17 (1965) 449–467.
- [33] H.W. Kuhn, Variants of the Hungarian method for assignment problems, *Nav. Res. Logist. Q.* 3 (4) (1956) 253–258.
- [34] J.W. May, *Cheminformatics for Genome-Scale Metabolic Reconstructions (Doctoral Dissertation)*, University of Cambridge, 2015.
- [35] N. Benchettara, R. Kanawati, C. Rouveirol, Supervised machine learning applied to link prediction in bipartite social networks, in: *2010 International Conference on Advances in Social Networks Analysis and Mining, IEEE, 2010, August*, pp. 326–330.
- [36] J.A. Storer, *An Introduction to Data Structures and Algorithms*, Springer Science & Business Media, 2012.
- [37] X. Feng, W. Wei, X. Li, Z. Zheng, Core influence mechanism on vertex-cover problem through leaf-removal-core breaking, *J. Stat. Mech. Theory Exp.* 2019 (7) (2019) 073401.
- [38] W. Wei, R. Zhang, B. Guo, Z. Zheng, Detecting the solution space of vertex cover by mutual determinations and backbones, *Phys. Rev. E* 86 (1) (2012) 016112.
- [39] A.R. Benson, D.F. Gleich, J. Leskovec, Higher-order organization of complex networks, *Science* 353 (6295) (2016) 163–166.
- [40] A. Sarajlić, N. Malod-Dognin, Ö.N. Yaveroğlu, N. Pržulj, Graphlet-based characterization of directed networks, *Sci. Rep.* 6 (2016) 35098.