

Article

A New Method for 3-Satisfiability Problem Solving Space Structure on Structural Entropy

Chen Liang ¹ , Xiaofeng Wang ^{1,2,*}, Lei Lu ¹ and Pengfei Niu ¹

¹ College of Computer Science and Engineering, North Minzu University, Yinchuan 750021, China; 20187166@stu.nmu.edu.cn (C.L.); 20197234@stu.nun.edu.cn (L.L.); 20197232@stu.nun.edu.cn (P.N.)

² Ningxia Key Laboratory of Information and Big Data Processing, North Minzu University, Yinchuan 750021, China

* Correspondence: xfwang@nmu.edu.cn; Tel.: +86-189-0958-7128

Abstract: Analyzing the solution space structure and evolution of 3-satisfiability (3-SAT) problem is an important way to study the difficulty of the solving satisfiability (SAT) problem. However, there is no unified analysis model for the spatial structure and evolution of solutions under different constraint densities. The analysis of different phase transition points and solution regions is based on different metric analysis models. The solution space of 3-SAT problem is obtained by planting strategy and belief propagation. According to the distribution of the influence of frozen variables on the solution, a label propagation algorithm based on planting strategy is proposed, is used to find the solution cluster, and then the structure entropy is used to measure its structure information. The structure entropy analysis model of 3-SAT problem solution space is established, and the unified analysis framework of solution space evolution and satisfiability phase transition is given. The experimental results show that the model is effective and can accurately analyze the evolution process of solution space and satisfiability phase transition, and verify the accuracy of interference phase transition point threshold predicted by long-range frustration theory.



Citation: Liang, C.; Wang, X.; Lu, L.; Niu, P. A New Method for 3-Satisfiability Problem Solving Space Structure on Structural Entropy. *Symmetry* **2021**, *13*, 2005. <https://doi.org/10.3390/sym13112005>

Academic Editor: Alexei Kanel-Belov

Received: 16 September 2021

Accepted: 19 October 2021

Published: 22 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: the satisfiability problem; structure entropy; belief propagation algorithm; label propagation algorithm

1. Introduction

The k -satisfiability (abbreviated) problem refers to a Boolean formula k given a conjunctive normal form (CNF), in which each basic clause is a disjunction of k words to determine whether it exists. A set of truth assignments makes the value of F true and finds all assignments that can satisfy F . The theory and application of the SAT problem is an important topic for scholars of computer and mathematical logic [1]. The k -satisfiability (k -SAT) problem is also the first problem to be proven to be non-deterministic polynomial(NP)-complete.

Since the discovery of phase transition phenomenon [2–4] in solving the random 3-SAT formula in the early 1990s, researchers have found that the random SAT formula is almost always satisfiable or unsatisfiable, depending on the size of the constraint density α . The change point is called the satisfiable phase change point α_s . In the random 3-SAT problem, although the exact value of α_s has not been studied, studies have shown that α_s is at least 3.52 and at most 4.506. Before α reaches α_s , the solution space of the 3-SAT formula undergoes several phase transitions and evolutions.

In recent years, scholars have conducted extensive research on the k -SAT problem. The most important progress includes the prediction of the phase transition point, the search for the strict lower bound of the satisfiability threshold, etc., and a wealth of research results have been obtained. Among them, literature [5,6] uses the symmetric cavity method (1RSB) in statistical physics to give the satisfiability threshold α_s of the 3-SAT formula as a function of k , and calculate the approximate value $\alpha_s \approx 4.267$. Literature [7–9] used

the 1RSB entropy cavity method [10,11] and found that before α reaches α_s , the solution space undergoes a cluster phase transition at $\alpha = \alpha_d$. The solution space before the phase transition is mainly dominated by a large solution cluster, in which there are rich internal structures and multi-dimensional communities. As α increases, when α reaches α_d , the number of macroscopic states in the solution space will suddenly increase, that is, the solution space will begin to evolve and split into exponentially multi-level declusters, and solutions of different declusters cannot be reversed by a single variable. From one cluster to another cluster, and its similarity is much lower than the internal solutions of the same cluster, it is more difficult to use the local search algorithm to solve. After the clustering phase transition occurs, α continues to increase, and the constrained density α undergoes a condensed phase transition at α_c . At this time, the exponential multilevel clusters decrease sharply, but the solution space still has exponential multilevel declustering at this time, and the declustering. The separation is obvious. In the 3-SAT problem $\alpha_d = \alpha_c$, literature [12,13] used long-range frustration theory to predict the threshold of interference transition α_j^∞ in the k -SAT formula. When α reaches this threshold, it indicates freezing. The argument begins to strongly dominate each solution community in the declustering, and increasing the α declustering will cause the community to disappear. This threshold can also be used as the lower bound of the satisfiability phase transition point α_s , but it is shown through experiments that the prediction is accurate when $k \geq 4$. When $k = 3$, $\alpha_j^\infty = 4.1897$, which is quite different from the predicted result.

In 2016, Angsheng Li proposed the structural entropy theory [14]. The structural entropy principle of graph is essentially a measure, which is defined as determining the minimum total number of digits required for node code that can be accessed from the random walk in graph. This principle can completely or maximally detect the two-dimensional and k -dimensional structure composed of the rules and order of the graph against the random changes in the graph. They use random walk to capture the information interaction between nodes in the graph to completely distinguish the structural noise in the undirected graph, and use the structural entropy of the graph to study the information interaction between nodes and the natural clustering of nodes, and achieve some new results [15–17].

As is known to all, the internal structure of the solution space of SAT problem exists rich and multistage community structure, the view of the evolution and phase transformation of solution space, the existing studies are all narrow discussions of α with different constraint densities, lacking a unified generalized model. In addition, the existing research lacks a high summary of multidimensional structure information in solution space. K -SAT problems can be transformed into 3-SAT problems through the method of specification [18], so 3-SAT problems are of great research significance. Inspired by structural information theory, in this paper, we study the structural properties of the solution space of 3-SAT problem by using the structural entropy of graphs.

This paper studies the 3-SAT problem and the community structure changes in the problem solution space. Based on structural entropy theory, this paper proposes a two-dimensional structural entropy measurement model and a k -dimensional solution space measurement model to measure and analyze the evolution and phase transition of the 3-SAT solution space. We introduce the $G = (n, 3, \alpha)$ model to randomly generate instances of the same scale with different α . Furthermore, we propose and use the belief propagation (BP)-based planting strategy algorithm and the planting strategy-based label propagation algorithm (LPA) to generate the 3-SAT problem and the solution space, and study its structural information. Combined with the phase transition of the solution space structural entropy, the phase transition of the solution space and the change in community structure dependent on the change in α are analyzed; the structural entropy conditions when the solution space changes in the community are obtained, which was experimentally verified [12,13]. As result, the accuracy of the predicted interference transition threshold α_j^∞ could be determined.

2. Preliminaries

2.1. Bp Iterative Equation

Let us suppose the CNF is $F = (C_1, C_2, \dots, C_m)$, the set of arguments is denoted by $\{x_1, x_2, \dots, x_n\}$, and i is used to represent x_i . Formula F can construct a bipartite graph $G = (C \cup X, E)$, termed a factor graph. The edges in figure G are divided into real edges and imaginary edges. Among these edges, is called the argument set and $C = \{C_1, C_2, \dots, C_m\}$ is called the clause set. The conjunction of clauses constitutes the conjunctive normal form formula, which is denoted as $F = (C_1 \wedge C_2 \wedge \dots \wedge C_m)$ and $F = (C_1, C_2, \dots, C_m)$, for short.

As shown in Figure 1, two types of information are defined on the edge of the factor graph, namely the information $\mu_{i \rightarrow a}(s_i)$ ($s_i \in \{0, 1\}$) transmitted by the argument to the clause and the information $n_{a \rightarrow i}(s_i)$ sent by the clause to the argument [19,20]. $\mu_{i \rightarrow a}(s_i)$ indicates whether the argument i is in or not, the probability that the value is s_i under the constraint of clause a , and the probability that the variable i takes the value of s_i without the constraint of clause a . $n_{a \rightarrow i}(s_i)$ represents the probability that the variable i takes the value of s_i to satisfy clause a . According to the cavity theory in statistical physics, the following BP iterative equation can be obtained [21]:

$$\mu_{i \rightarrow a}^{(t)}(s_i) = \frac{\prod_{b \in (\partial i \setminus a)} n_{b \rightarrow i}^{(t)}(s_i)}{Z^{i \rightarrow a}} \tag{1}$$

$$n_{a \rightarrow i}^{(t+1)}(s_i) = \frac{\sum_{j \in (\partial a \setminus i), s_j \in D} \delta_{(s_i, s_j) \in Q_a} \mu_{j \rightarrow a}^{(t)}(s_j)}{Z^{a \rightarrow i}} \tag{2}$$

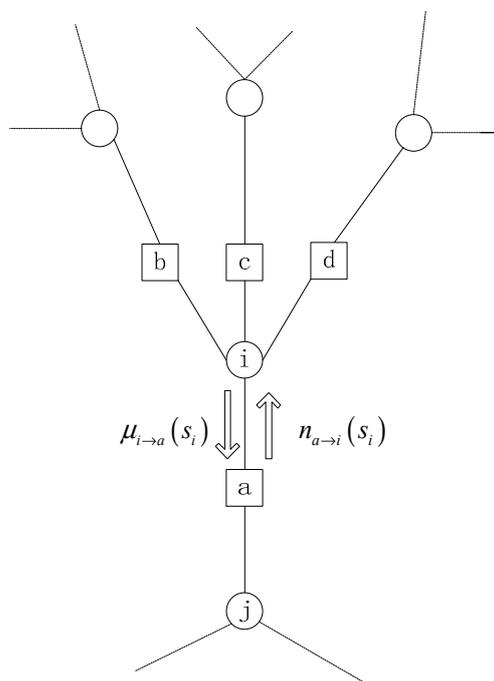


Figure 1. Factor diagram representation of BP iterative process on a random instance.

Among these formulas, $Z^{i \rightarrow a}$ and $Z^{a \rightarrow i}$ are normalization factors, $b \in (\partial i \setminus a)$ and $j \in (\partial a \setminus i)$, respectively, represent the set of clauses connected to the argument i to remove a and the set of argument nodes connected to the constraint a to remove i . If the BP iteration can converge, the marginal probability of argument i can be calculated according to the information $n_{a \rightarrow i}(s_i)$, and the value of the argument can be obtained from the marginal probability.

2.2. Solution Space of 3-SAT Problem

In the 1RSB theory, a solution cluster in the solution space of a random 3-SAT problem is regarded as a macroscopic state of the system. When the constraint density of the random 3-CNF reaches a certain critical value α_d , the structure of the solution space of the random k -CNF formula will change qualitatively, and the number of macro-states in the solution space will suddenly increase, that is, the solution space is split into a large number of subspaces, each subspace contains a certain number of solutions, but the statistical properties of different subspaces are different, and the similarity of the solutions of different subspaces is much lower than that of the same subspace. This abrupt phenomenon is called the cluster phase transition under the thermodynamic limit $N \rightarrow \infty$, and α_d is called the confinement density of the cluster phase transition.

Take the change of solution space of k -SAT formula as an example, as shown in Figure 2. When the constraint density α gradually increases from zero and is close to α_d , although the solution space has only one macroscopic state under the meaning of statistical significance, more and more community structures and more and more community structures have been accumulated in this macroscopic state. Strong point-to-set association. The point-to-collection feature correlation length tends to be infinite as $\alpha_d - \alpha$ becomes smaller. In the random k -SAT problem, when α exceeds α_d until it reaches another critical value point α_c , the solution space of the random k -CNF formula will be condensed. Regarding [22], it signifies that the statistical properties of the solution space begin to be determined by those subspaces that contain the largest number of microscopic configurations but a small number. In the random 3-SAT problem, $\alpha_d = \alpha_c$, that is, the cluster phase transition point and the condensed phase transition point coincide. When $\alpha = \alpha_d$, the number of subspaces that determine the statistical properties of the solution space changes from the order of $O(e^{cN})$ to the order of $O(N^b)$ (where c and b are constants).

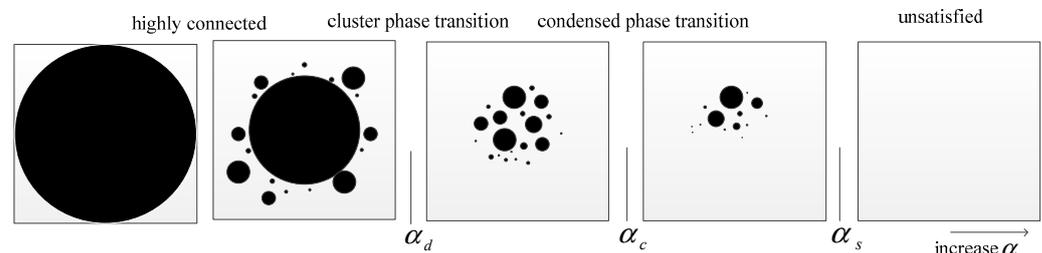


Figure 2. The solution space of the random k -SAT formula increases with α .

2.3. Two-Dimensional Structural Entropy of Undirected Graphs

The solution space is expressed as an undirected graph, which is convenient for us to study the structural characteristics of the solution space. The real-world network is a highly connected graph, which develops dynamically. We use random walks to obtain the interactive information of nodes in the network, and give the concept of entropy reflecting the dynamic complexity of the network. This is the one-dimensional structural information of the network. Here we first recall the Shannon entropy, and then introduce the definition of structural entropy [14].

Definition 1. For the probability vector $\mathbf{p} = (p_1, \dots, p_n)$ and $\sum_{i=1}^n p_i = 1$, then the entropy function of the probability vector \mathbf{p} is defined as:

$$H(p_1, \dots, p_n) = -\sum_{i=1}^n p_i \log_2 p_i \quad (3)$$

We interpret $-\log p_i$ as “ p_i 's self-information”, which also means that $-\log p_i$ is the amount of information required by a certain code. Therefore, $-\sum_{i=1}^n p_i \log_2 p_i$ is the average amount of information in the code $\mathbf{p} = (p_1, p_2, \dots, p_n)$ that determines the probability distribution.

Definition 2. Let $G = (V, E)$ be an undirected connected graph with n nodes and m edges. For each node $i \in \{1, 2, \dots, n\}$, d_i is the degree of node i in graph G , let $p_i = \frac{d_i}{2m}$, and then describe the fixed distribution of random walk through probability vector $\mathbf{p} = (p_1, p_2, \dots, p_n)$, then the definition of one-dimensional structural entropy is as follows:

$$H^1(G) = H(\mathbf{p}) = H\left(\frac{d_1}{2m}, \dots, \frac{d_n}{2m}\right) = -\sum_{i=1}^n \frac{d_i}{2m} \cdot \log_2 \frac{d_i}{2m} \quad (4)$$

The one-dimensional structure information $\mathcal{H}^1(G)$ of the connected graph measures the amount of information needed to determine the one-dimensional code of the node. The one-dimensional code can be obtained from a random walk in G and has a stable distribution.

Definition 3. For the undirected graph $G = (V, E)$, assuming that $P = \{X_1, X_2, \dots, X_L\}$ is the community of the node set V of G , the structural entropy of graph G with respect to the partition P is:

$$\begin{aligned} H^P(G) &= \sum_{i=1}^L \frac{V_i}{2m} \cdot H\left[\frac{d_1^{(i)}}{V_i}, \dots, \frac{d_{n_i}^{(i)}}{V_i}\right] - \sum_{i=1}^L \frac{g_i}{2m} \log_2 \frac{V_i}{2m} \\ &= -\sum_{i=1}^L \frac{V_i}{2m} \sum_{i=1}^{n_i} \frac{d_i^{(i)}}{V_i} \log_2 \frac{d_i^{(i)}}{V_i} - \sum_{i=1}^L \frac{g_i}{2m} \log_2 \frac{V_i}{2m} \end{aligned} \quad (5)$$

where n_i represents the number of nodes in community X_i ; V_i is the volume of community X_i ; that is, the sum of degrees in community X_i ; L represents the number of communities in partition P ; $d_j^{(i)}$ is the degree of the i -th node in X_i ; and g_i is community X_i , which connects degrees outside the community. If each node in graph G is regarded as a code pair (i, j) , the code of the community-containing node v is i , and the code of this node in the community is j . According to the information theory proposed by Shannon, the first term of $H^P(G)$ is the minimum number of coding bits required to determine the position of node v in community X_i , that is, the node that graph G can visit through one-step random walk is v . The second item is to determine the minimum number of coding bits required for community X_i . Therefore, the two-dimensional structural entropy is defined as:

$$H^2(G) = \min_P \{H^P(G)\} \quad (6)$$

Definition 4. For the two-dimensional structural information of a discontinuous graph, given a graph G , suppose it is an inductive subgraph of all connected components of G_1, G_2, \dots, G_L , and there is a two-dimensional structural entropy formula:

$$H^2(G) = \frac{\sum_{i=1}^L \text{Vol}(G_i) \cdot H^2(G_i)}{\text{Vol}(G)} \quad (7)$$

where $\text{Vol}(G)$ is the volume of the graph G , $\text{Vol}(G) = \sum_{g \in G} d_g$ is defined according to the two-dimensional structural entropy, and G is the weighted average of the two-dimensional structural entropy of all subgraphs G_i . Specifically, if there is a subgraph, such as G_j , with only one isolated node, the structural information of G_j is: $\lim_{p \rightarrow 0} -p \log_2 p = 0$. Therefore, for any graph G , if the G node has no edges, then $H^2(G) = 0$.

Information entropy reflects the overall amount of information in the undirected graph, that is, the uncertainty of the undirected graph network. Based on this, the two-dimensional structural entropy follows the principle of minimizing uncertainty to remove structural noise and data redundancy. The degree of association between structure and node structure can effectively measure the structural information within a community and the structural complexity between different communities.

3. Structural Entropy Measurement Model of 3-SAT Formula

3.1. Algorithm of Planting Strategy Based on BP

Since the purpose of establishing the solution space is to study its internal structure, it is necessary to solve a large 3-CNF. The 3-SAT problem is an NP complete problem, and it is difficult to use a complete algorithm to find all its solutions. The information dissemination algorithm is currently the most effective method to solve the 3-SAT problem. For example, the belief dissemination algorithm can effectively solve the difficult areas of $\alpha_d < \alpha < \alpha_c$, and the information dissemination process can also efficiently find the frozen variables. It is suitable for planting strategies to find α located in a large number of solutions in the easy solution area of $\alpha < \alpha_d$.

The planting solution strategy [8,23–25] starts by planting the frozen variable in the solution group, and finds the solution clustering in the solution group by randomly flipping the non-frozen variables through random walks. In order to make the algorithm reach a fixed point, we start from the external nodes to fix the variables. We then fix some of the variable nodes through the BP iterative equation, expand to decluster through a series of unbiased random walks, and then generate the declustering with flipped labels. If we let the assigned $S \equiv \{S^1, S^2, \dots, S^N\}$ satisfy formula F by a set of solutions, the Hamming distance between the two solutions is defined as:

$$d(S_1, S_2) = \sum_{j=1}^N \delta(S_1^j, -S_2^j) \quad (8)$$

The adjacency matrix is constructed by the Hamming distance between the solutions, and then an undirected graph of the solution space is generated. We classify the arguments:

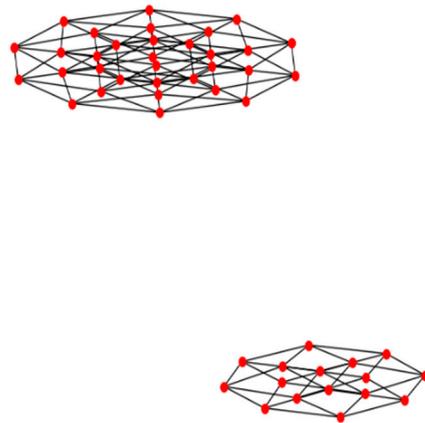
Type A—variables that have been fixedly assigned and;

Type B—variables without fixed assignments.

In Algorithm 1, Steps 1–7 use BP iteration to calculate the satisfiable probability of each variable and divide the variable nodes into two categories, namely, frozen variables and non-frozen variables. In Step 8, the non-frozen variables are randomly flipped by the walkSAT algorithm to form a solution space with the frozen variables as the core, and the number of flips for generating each solution walkSAT is recorded. Steps 9–10 calculate the Hamming distance between the solutions and generate the solution space adjacency matrix to construct the solution space. The solution space obtained through the planting strategy records the number of flips of each solution, which can be used as the basis for the next solution space community division. The number of flips means the influence of the frozen variable on the solution. The more the number of flips, the greater the influence, and the more likely it is to be in the same community. The solution space generated by Algorithm 1 is shown in Figure 3.

Algorithm 1 Algorithm of planting strategy based on BP.**Input:** CNF F , maximum number of iterations t_{max} , convergence accuracy ϵ , fixed argument accuracy θ **Output:** Solution space undirected graph

1. For each edge (i, a) , randomly initialize the constraint to the variable information $n_{a \rightarrow i}^{(0)}(s_i) \in [0, 1]$.
2. For each edge (i, a) , update the information according to Equations (2) and (1).
- if $\partial i \setminus a = \emptyset$, let $\mu_{i \rightarrow a}^{(t-1)}(s_i) = \frac{1}{d_N}$. Then, enter the value of $\mu_{i \rightarrow a}^{(t-1)}(s_i)$ into Equation (2) to obtain $n_{a \rightarrow i}^{(t)}(s_i)$.
3. **if** $|n_{a \rightarrow i}^{(t)}(s_i) - n_{a \rightarrow i}^{(t-1)}(s_i)| < \epsilon$, then $n_{a \rightarrow i}^*(s_i) = n_{a \rightarrow i}^{(t)}(s_i)$. Go to Step 5.
- if** $t > t_{max}$, t is marked as a Type B variable.
4. Go to Step 2.
5. Calculate the marginal probability $n_i(s_i) = \frac{\prod_{a \in \partial i} n_{a \rightarrow i}^*(s_i)}{\sum_{s_i \in D} \prod_{a \in \partial i} n_{a \rightarrow i}^*(s_i)}$ of each variable according to the following formula.
6. **if** $t > t_{max}$, select variables for assignment. For each variable:
if $S_i < \theta$, $S_i = -\sum_{s_i} n_i(s_i) \log(n_i(s_i))$
if $n_i(1) > n_i(0)$, n_i is marked as fixed argument (Type A), $i = 1$;
else if $n_i(1) < n_i(0)$: n_i is marked as a Type A as i is a fixed argument;
else: i is marked as a Type B argument.
7. Traverse the CNF formula argument, if Argument A: delete the argument, update CNF formula F ; **else**: go to Step 8.
8. Import the updated formula into the walkSAT solution and record the number of flips S^{flip} of each solution.
9. Use Formula (8) to calculate the distance between the solutions and construct the adjacency matrix
if $d(S_i, S_j) = 1$, continue;
else $d(S_i, S_j) = 0$.
10. Establish an undirected graph of the solution space according to the adjacency matrix;

**Figure 3.** The solution space generated based on a planting BP strategy algorithm ($\alpha = 4.10$, $n = 100$).

3.2. K-Dimensional Structural Entropy of Undirected Graphs

Definition 5. For the undirected connected graph $G = (V, E)$, the multilayer partition tree T of G is set by the method described in Section 3, and the undirected connected Graph G is divided into k layers. If node α is a tree node and α is not the root node, then the definition of k -dimensional structural entropy is:

$$H^T(G) = \sum_{\alpha \in T, \alpha \neq \lambda} H^T(G, \alpha), \quad (9)$$

$$H^T(G, \alpha) = -\frac{g_\alpha}{2m} \log_2 \frac{V_\alpha}{V_\alpha -}$$

Among them, $m = |E|$; g_α is the number of edges of the nodes of the community T_α connected to nodes outside the community and V_β is the cumulative sum of the T_β degree of the set T_β . The k -dimensional structural entropy of G is defined as follows:

$$H^K(G) = \min_{T: H(T) \leq K} \{H^T(G)\} \quad (10)$$

It can be seen from the above formula that an algorithm that approximates the k -dimensional structural information of the graph by finding the k -dimensional structure that minimizes the entropy of this value can greatly reduce the complexity of the multidimensional structural information of the graph. By finding an optimal coding tree $T(H(T) \leq K)$, it is possible to make $H^{K+1} \leq H^{T^*}(G)$. Therefore, the k -dimensional structural entropy of G has another definition: $H^K(G) = H^T(G)$.

4. Structural Metric Model of Solution Space of 3-SAT Formula

4.1. Lpa Based on Planting Strategy

The definition of structural entropy involves the concept of community division. This means a complex network is divided into several non-overlapping communities according to a certain method; the structural information within and between communities is then calculated through classification and integration, and the structure of the entire network is finally obtained. Entropy is a manifestation of important structural information, such as the structural complexity, dynamic interaction, and uncertainty of the network.

There are no specific requirements for community division in the definition of structural entropy, but according to Formula (5), improving the accuracy of community division is the only way to minimize network uncertainty. Currently, in broad sense, there is no efficient community discovery algorithm that can satisfy any community at any same time. Algorithms that perform well in individual networks may not perform satisfactorily in community discovery in other networks. Therefore, designing an efficient community discovery algorithm for the solution space of the 3-SAT problem can greatly improve the accuracy of the structural entropy measurement method of the solution space in this problem.

The LPA is a semi-supervised machine-learning method based on graphs. It has the advantage of being suitable for large-scale undirected graphs; its community discovery is fast and the number of communities does not need to be known before each iteration.

When the node is initialized, the original LPA assigns a unique label to each node in the network, which randomly arranges the nodes to obtain the node sequence and uses this as the label update sequence. In the calculation, if the highest number of labels of neighboring nodes is the same, then a neighbor is randomly selected to update the label. When using the LPA algorithm for community discovery, the update order of tags is very sensitive; thus, the update order of these tags affects the magnitude of community discovery. The above two random processes will gradually increase the impact of relatively important nodes on the community as the iteration progresses, while ignoring the differences in importance between nodes, resulting in a "backflow" of labels, and may cause problems that did not originally belong to the community. Nodes are included in the community, which ultimately lead to too much difference between the results discovered by the community and the real community.

In the solution space generated by Algorithm 1 in the previous section, an LPA based on the planting solution strategy is proposed to optimize the belief propagation algorithm to solve the community division of the solution space in the 3-SAT formula.

In the process of selecting the label update, the initially labeled node selects the label with the highest occurrence in the neighbor nodes to update the label, and selects the label to be updated according to Equation (11):

$$l_i = \operatorname{argmax} \sum_{j \in N(i)} \delta(l_j, l) \quad (11)$$

$$\delta_{ij} = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$$

where l_i represents the label to be updated, $N(i)$ represents the set of neighbor nodes of node i , and l_j represents the label of node i 's neighbor node j .

In the solution space calculated based on the planting strategy, the greater the number of common neighbors between nodes, the greater the number the nodes that are affected by the same or several frozen variables. At the same time, each node has a parameter of the

number of flips. The smaller the number of flips, the more the node is affected by the frozen variables. As a gradually increases, the influence of the frozen variable on the solution space community gradually increases, so the community must be divided according to the influence of the frozen variable on the solution. Dividing the community structure in this way is also more accurate than the traditional LPA and can greatly reduce the uncertainty of the solution space structural information.

The LPA based on the planting solution strategy proposed in this paper uses the influence of the above frozen variables to determine the update order. It also uses the similarity between nodes to measure the more common neighbors of the two nodes: the greater the similarity of the nodes, the greater the probability of being in the same community, thus the result of the division is more in line with the original community distribution. The degree of a node in an undirected graph is defined as the number of edges directly connected to the node. The connected nodes structurally indicate that there is only one variable value between them. The similarity between nodes is calculated by the following formula:

$$S(v_i, v_j) = \frac{|NG(v_i) \cap NG(v_j)|}{d_{v_i}} \quad (12)$$

Among them, $NG(v_i)$ is the first-order neighbor set of nodes v_i and d_{v_i} is the degree of node v_i . In this algorithm (Algorithm 2), the node update sequence is traversed, and the similarity calculation is performed. If the current node and the node before the update sequence are adjacent nodes, the similarity index is less than the threshold. If the label of the node is null, then the node before the sequence is selected as the seed node; otherwise, it is not selected as the seed node.

Algorithm 2 LPA based on planting solution strategy.

Input: Tagged network $G = (V, E)$, threshold λ

Output: Community findings

1. **If** $G(A)$ connect: go to Step 2, flag = 1;
else: Backtrack and traverse all connected subgraphs; go to Step 2, flag = 0.
 2. According to the walkSAT flip result, the ascending sequence $A = \{i_1, i_2, \dots, i_n\}$ of the number of nodes set flips is formed, and the seed node set $S = \{\}$ is generate.
 3. $t = 1$ for $1 \leq j \leq n$, let $L(i_j) = \{\}$.
 4. $t = t + 1, j = 1$; if $t > n$: go to Step 6.
 5. $j = j + 1$, if $j \geq t$: go to Step 4.
 - (1) Traverse the connected graph nodes and calculate the similarity between the nodes according to Formula (10), if $s(i_t, i_j) < \lambda$, go to Step 5.
 - (2) **If** $i_j \in S, L(i_t) \cup \{j\}$: go to Step 5.
 - (3) **If** $i_j \in S, L(i_j) = \{\}, S = S \cup \{i_j\}, L(i_j) = \{j\}$, and $L(i_t) = L\{i_t\} \cup \{j\}$: go to Step 5.
 - (4) **If** $i_j \in S, L(i_j) \neq \{\}$, thus $L(i_t) = L\{i_t\} \cup \{j\}$: go to Step 5.
 - (5) Go to Step 5
 6. **If** there are multiple labels for node v in the network, the label with the most occurrences in its neighborhood is selected as the label for v .
 7. **If** there are multiple seed nodes that meet the conditions, the one with the most frequent occurrences of the label is selected as its own label. If node v does not have an adjacent seed node that meets the conditions, keep the label as empty v .
 8. Establish and traverse the node set \bar{V} with empty labels in the current network, select the label with the most occurrences in neighboring nodes as the label of v and retrospectively update to \bar{V} without change.
-

In the above algorithm, Steps 1–7 use BP iteration to calculate the satisfiable probability of each variable and divide the variable nodes into two categories, namely, frozen variables and non-frozen variables. In Step 8, the non-frozen variables are randomly flipped by the walkSAT algorithm to form a solution space with the frozen variables as the core, and the number of flips for generating each solution walkSAT is recorded. Steps 9–10 calculate the

Hamming distance between the solutions and generate the solution space adjacency matrix to construct the solution space.

The number of flips means the influence of the frozen variable on the solution. The more the number of flips, the greater the influence, and the more likely it is to be in the same community. The solution space obtained by Algorithm 1 through the planting strategy records the number of flips of each solution, which can be used as the basis for the next solution space community division.

4.2. Two-Dimensional Structural Entropy of Solution Space of 3-SAT Formula

In order to obtain the structural information of the solution space of the 3-SAT formula, the solution is represented by nodes, and the connection between nodes means that only one argument between the two solutions has a different value. This can be completed without any loss of the relationship between the solutions and the relationship between the solutions and the information of each node. The solution generated by randomly flipping the non-frozen variables on the basis of the frozen variables by walkSAT in Step 8 of Algorithm 1 is abstracted as an undirected graph. Therefore, through the research and analysis of the solution space undirected graph, the structure information of the solution space of the 3-SAT problem can be obtained more directly. The community structure of the undirected graph divided by Algorithm 2 is shown in the Figure 4.

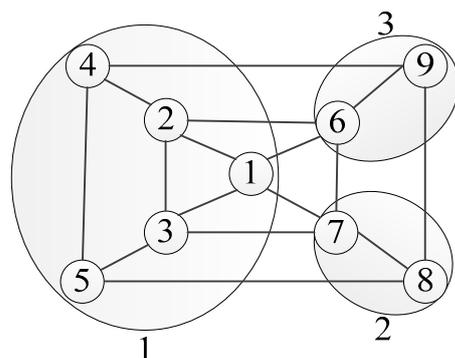


Figure 4. Algorithm 2 iteratively divides the solution space subcommunity process.

The nodes in Figure 3 are divided into three communities, namely {1,2,3,4,5}, {7,8}, {6,9}. The division of the community structure not only depends on the inherent relationship between nodes and more importantly, divide the nodes with similar influence distribution of the frozen variables on the solution community into the same community, and always follows the principle of minimizing structural entropy in the process of radiating other nodes. Therefore, the community structure can be maximized. It satisfies the information minimization principle in the definition of structural entropy, that is, it satisfies $H^2(G) = \min_p \{H^p(G)\}$ in the definition of two-dimensional structure entropy and $H^K(G) = \min_{T:H(T) \leq K} \{H^T(G)\}$ in the definition of K-dimensional structure entropy.

If the undirected graph G has no self-loops and polygons, then the probability of a random walk in G is set $p_t = \frac{d_t^{(p)}}{V_j}$, the static distribution of the random walk can be expressed as a probability vector $p = (p_1, p_2, \dots, p_n)$, where $\sum_{i=1}^n p_i = 1$. The amount of information about the distribution can be obtained:

$$I(G) = I(p) = -\log_2 \frac{d_i^{(j)}}{V_j} \tag{13}$$

According to the theory of information entropy, in a limited set of mutually exclusive and joint exhaustive events, entropy is the average amount of information, so the local

structural entropy (average amount of information in the network structure) within the community is expressed as:

$$H^{\text{inside}}(G) = - \sum_{j=1}^L \frac{V_j}{\text{vol}(G)} \sum_{i=1}^{n_j} \frac{d_i^{(j)}}{V_j} \log_2 \frac{d_i^{(j)}}{V_j} \quad (14)$$

In information theory, log represents the unit of entropy as “bit”, that is, the entropy value represents the number of bits required for encoding. Therefore, $-\sum_{i=1}^q p_i \log_2 p_i$ represents the average amount of information required to determine the number of coded bits of i by the probability $p = (p_1, p_2, \dots, p_n)$, where $l = -\log_D p_i$ is used as the number of binary representations, indicating that $\frac{1}{p_i}$ can be represented as D^l , so $-\log_2 p_i$ is to determine the amount of information needed to encode i . The local structural entropy between communities is expressed as:

$$H^{\text{outside}}(G) = - \sum_{j=1}^L \frac{g_j}{\text{vol}(G)} \cdot \log_2 \frac{V_j}{\text{vol}(G)} \quad (15)$$

The community structure is divided by the LPA algorithm based on the planting strategy, and the sum of Formulas (14) and (15) are used as the two-dimensional structural entropy of the proposition formula, which reflects the interactivity between nodes within the community. In addition, due to the consideration of the influence of frozen variables, these influences are used to form communities. According to the probability of random walk and the random flip of walkSAT, the understanding nodes are obtained by radiation expansion. At the same time, it is changed to Algorithm 2 and it can extract structural noise and data redundancy effectively, which satisfies the principle of minimizing the uncertainty in the undirected graph community discovery to the greatest extent.

4.3. K -Dimensional Structure Entropy of Solution Space of 3-Sat Formula

The previous section has analyzed that the LPA based on planting strategy follows the principle of minimizing structural entropy when dividing communities, and satisfies $H^K(G) = \min_{T:H(T) \leq K} \{H^T(G)\}$ in the definition of k -dimensional structural entropy. This section introduces the construction process of the K -dimensional structural entropy community segmentation tree.

As shown in Figure 5, using Algorithm 2 iterative community division, the solution space undirected graph can be divided into a multi-level community structure. As shown in Figure 6, thereby constructing a split tree with a set of nodes in the graph as tree nodes. The tree structure can be constructed according to the following regulations:

- (1) Define the set $T_\lambda = V$, λ is the root node.
- (2) Let α be the node on the split tree. The immediate successor node of α is α_j , $j = \{1, 2, \dots, n\}$ and the nodes are sorted from left to right in ascending order.
- (3) Let α, β be the node on the split tree. $\alpha \subset \beta$ indicates that α is the initial character segment of β . When α is not the root node, the longest initial character segment of α is α^- .
- (4) For each node, $\{T_\alpha | h(\alpha) = i\}$ is a division of V ; thus ($h(\alpha)$ is the height of α , the height of the root node is 0, and the height of other nodes increases according to the tree).
- (5) For each node, if α is the parent node of β , $T_\alpha = \cup_{\beta^- = \alpha} T_\beta$.

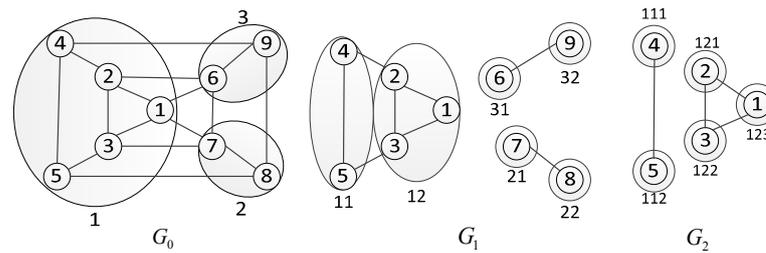


Figure 5. Solution Space K-Dimensional Community Division.

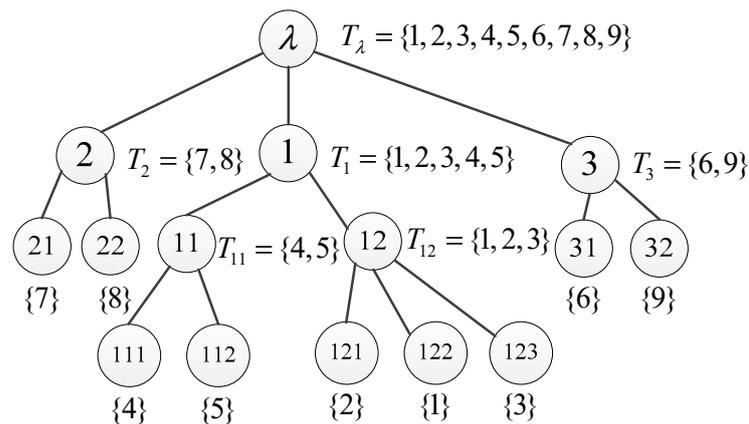


Figure 6. K-Dimensional Structure Entropy Partition Tree Construction.

After initializing the split tree, the undirected graph is divided into communities to obtain multiple overlapping communities M_1, M_2, \dots, M_n , and these overlapping communities are used as the second-level nodes of the split tree, where the parent nodes are the root nodes λ (including all nodes in the undirected graph). Each community contains one or more nodes, and $G = M_1 \cup M_2 \cup \dots \cup M_n$. If the number of nodes in M_i ($i \in n$) is greater than one, continue to divide M_j , and use the sub-community $M_{i1}, M_{i2}, \dots, M_{ij}$ as its child nodes. If the sub-community still contains multiple nodes, repeat the above steps to continue dividing the community until it is in the community. When only a single node is included, the division of the community is stopped.

For a given undirected graph G , according to the principle of information minimization, the K -dimensional structure information of G is to determine the total number of bits of K -dimensional coding of G . Therefore, the K -dimensional structural entropy also provides the dynamic complexity of the quantized undirected graph G of structural information. On the one hand, as the constraint density α reaches the satisfiability threshold α_s , a single solution cluster in the solution space of the 3-SAT formula undergoes a cluster phase transition [7–9], and the cluster phase transition before the phase transition has rich internal structure, and most of the solutions in the solution space are dominated by a large solution cluster. On the other hand, it is defined by the K -dimensional structure entropy. The K -dimensional knot structure entropy is suitable for the multi-level structure analysis of undirected connected graphs. Therefore, The K -dimensional structure entropy is suitable for measuring the multi-dimensional structure information of large-scale declustering before the cluster phase transition occurs.

It can be seen from Formula (8) that all the communities of the connected undirected graph are included in the K -dimensional structure entropy calculation based on the structure of the partition tree, and the rules and order of the K -dimensional structure graph can be detected to the greatest extent. The order is completely extracted from the undirected graph, and the order and disorder are separated from the structured data, and the cumulative sum of the entropy of the tree nodes is calculated, which conforms to the additivity of entropy. Therefore, the K -dimensional structure entropy takes into account the

community segmentation of all nodes, and fully and effectively analyzes and understands the dynamic complexity and uncertainty of spatial undirected graphs and other important properties. The partition tree constructed by the iterative division of Algorithm 2 can effectively remove the noise in the undirected graph, and minimize the information loss from the construction of the partition tree to the calculation process of structural entropy. The final calculation result is a highly abstract and summary of the K -dimensional information of declustering.

5. Result Analysis

In order to ensure the accuracy of the 3-SAT formula solution space measurement model, this section first verifies the performance of the LPA algorithm based on the planting strategy. For the communities divided by the BP algorithm (Algorithm 1) based on the planting strategy, the modularity standard is used to analyze the quality of the community division. The basic principle is to regard the network as a zero-model network, and discover the difference between the modularity of the community and the modularity of the zero model through comparison. The greater the difference, the better the quality of community discovery. The definition of modularity Q is:

$$Q = \frac{1}{2M} \sum_{ij} \left[a_{ij} - \frac{k_i k_j}{2M} \right] \delta(c_i, c_j) \quad (16)$$

where M represents the number of edges in the network, a_{ij} represents the adjacency matrix of the network, k_i represents the degree of node i , and the function $\delta(k_i, k_j) = 1$ represents if nodes i and j are classified as the same community. Because other community discovery algorithms (including LPA) cannot meet the pertinence of the model, a large number of experiments are of little significance, and Algorithm 2 has similar properties to LPA, so only the modularity of Algorithm 2 and LPA is compared. The comparison results of the 3-SAT formula under different constraint densities α are shown in Table 1.

Table 1. Algorithm 2 and LPA Modularity Comparison.

	F_1	F_2	F_3
LPA	0.3532	0.3296	0.3620
Algorithm 2	0.4236	0.4159	0.4571

Table 1 shows that in the solution space of the 3-SAT formula with different constraint densities, the modularity of the LPA based on the planting strategy is significantly improved than that of the LPA, and because Algorithm 2 is specific to the solution space community division, it shows that Algorithm 2 is more effective to solve the division of space community.

The hardware environment for the experiment is Intel Core i7 9750H + RTX 2060 + 16 GB RAM, and the software environment is Windows 10 \times 64 + ubuntu18.04 + xshell + Python3.7 + Matlab. Let us use the $G(n, m, \alpha)$ model to generate the 3-SAT formula, where n is the number of arguments and m is the number of clauses. We let $n = 100, 125, 150$, then generate a CNF formula with a length of 100, 125, 150, and increase the value of the constraint density by adding randomly generated clauses one by one. In the solution space undirected graph, Algorithm 2 is used to divide the communities, and the two-dimensional structural entropy of the solution space of the 3-SAT formula is calculated. The result is as follows:

Figure 7 shows the change of the entropy of the two-dimensional structure when the variable scale is 100 and the constraint density increases from 3.5 to 4.3. As the constraint density increases, the number of solutions and the complexity of the solution space begin to change, and the entropy of the two-dimensional structure decreases as α increases. When the constraint density is between 3.50 and 3.74, a single decluster has a complex community structure and the decrease in the entropy of the two-dimensional structure

$\Delta H^2(G)$ is not stable enough. As the number of solutions continues to decrease, the value of $\Delta H^2(G)$ generally shows a downward trend. As the constraint density increases, the constraint density undergoes a clustering phase transition at 3.73, and the declustering undergoes a heterogeneous transformation. A single decluster begins to decompose into multiple declusters, but the solution space is still dominated by one main decluster, here. In the process, the complexity of the internal declustering community continues to decrease, and the decrease in the constraint density around 3.92 tends to be flat. As the value of α increases to about 4.17, the solution space changes again, and the number of solutions continues to decrease. The frozen variable increases the impact on the solution community, the single cluster is sharply separated into smaller clusters, and the solution space begins. This solution space tends to be stable; and after increasing α , the solutions in the solution space begin to disappear in clusters. It can be seen from Figure 7 that the two-dimensional structural entropy can effectively measure the overall structural information of the solution space.

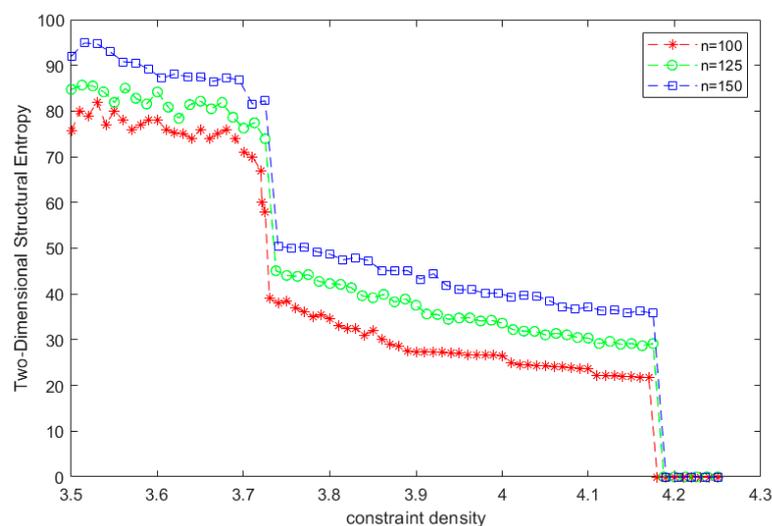


Figure 7. Two-Dimensional Structural Entropy Distribution.

The literature uses symmetric cavity theory (1RSB) [5,6] to provide a 3-SAT formula cluster phase transition point α_d around 3.859, which is lower than the first phase transition result in Figure 7. When α is near the first phase transition point, the complexity of solution space increases sharply, which leads to a sharp increase in structural information and a decrease in accuracy of Algorithm 2, so $H^P(G) < H^2(G)$, $H^T(G) < H^K(G)$. This leads to the experimental results having lower upper bounds than those in the literature [5,6]. When CNF is more easily divided, or more accurate algorithms are used, the upper bound of experimental results will be improved, but it cannot give a definite upper bound. These studies suggest that the condensed phase transition point in 3-SAT is $\alpha_c - \alpha_d$, that is, the few largest solution clusters that contain the most solutions in the solution space of the 3-SAT formula cannot determine the statistical properties of the solution space. Figure 7 shows that the second transition occurs when the constraint density is about 4.17, which is close to the use of remote frustration theory to predict the threshold α_j^∞ in the literature [12,13].

After the transition, the entropy of the two-dimensional structure tends to be segmented and stable. If the value of the frozen argument is large, it cannot satisfy the newly added random clause. As α increases, the declustering of the solution space begins to cluster and disappear, and the entropy of the two-dimensional structure also decreases in segments. This means that at this time the frozen variables have greatly affected every solution community, so it can be considered that this transition point is the interference transition point predicted by the remote frustration theory.

The above results show that the two-dimensional structural entropy accurately reflects the overall structural information of the solution space of the 3-SAT formula. The phase

transition and the transition point of the two-dimensional structural entropy illustrate the validity of this measurement.

Before the clustering phase transition occurs, each solution cluster in the solution space is relatively large, and the number of solutions has not decreased sharply. It contains complex community structures and rich multidimensional communities. The entropy of the two-dimensional structure cannot be accurately determined. The solution space structure is measured, that is, the decrease in the entropy of the $3.50 < \alpha < 3.73$ two-dimensional structure shown in Figure 7 is unstable. Therefore, this section uses the k -dimensional structural entropy to measure the k -dimensional structural information of a single decluster of the solution space. The solution space undirected graph data constructed in the process of calculating the entropy of the two-dimensional structure are used continuously, and Algorithm 2 is used to iteratively partition the community, divide the partition tree, and then calculate the k -dimensional structure entropy. The result is shown in Figure 8.

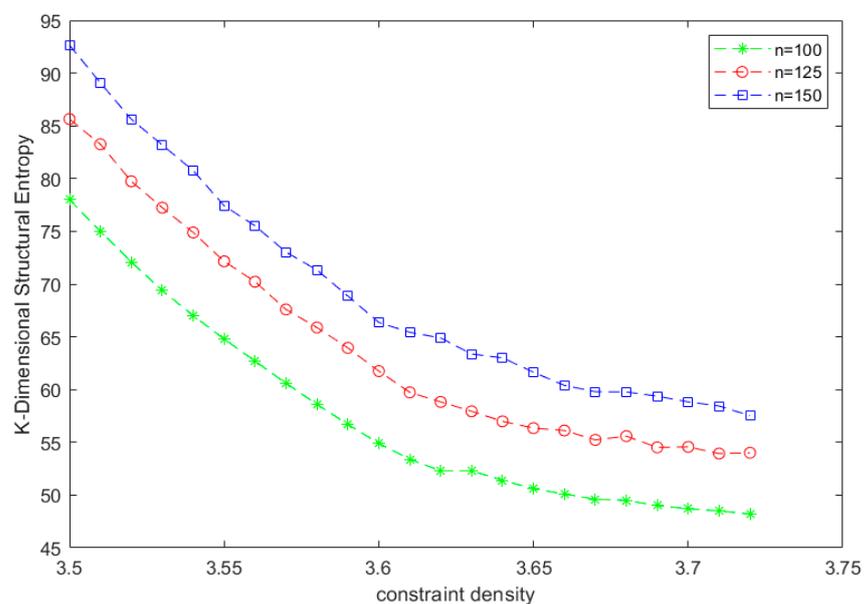


Figure 8. K-Dimensional Structural Entropy Distribution Before Single Cluster Decomposition.

Figure 8 shows that before the cluster phase transition occurs, the internal k -dimensional structural information of the solution space continues to decrease, and the k -dimensional structural entropy decreases steadily. At $\alpha = 3.62$, the decline decreases and gradually stabilizes. At $\alpha = 3.74$, a clustering phase change occurs, and the solution space that was originally a single declustering is decomposed into multiple declusterings. This phenomenon shows that the k -dimensional structural entropy performs well when measuring the solution space of single clustering, and is suitable for clustering with rich internal structure.

The above experimental results show that the 3-SAT formula can be used to solve the spatial clustering phase transition point α_d , the condensed phase transition point α_c , the threshold value of the interference transition point, and the corresponding structural entropy threshold value. At $n = 100$, if $H^K < 48.20$, then there is no cluster phase transition in the solution space. If $H^K > 48.20$, the solution space does undergo a cluster phase transition: the cluster phase transition point is around $\alpha_d = 3.73$, the condensed phase transition point $\alpha_c = \alpha_d$, the two-dimensional structural entropy H^2 has an interference transition at 24.2, and the interference transition point $\alpha_j^\infty = 4.17$. The above structural entropy analysis can obtain the structural entropy thresholds of the 3-SAT formula clustering phase transition points, the condensed phase transition points, and the interference transition points at different scales. It can also estimate the clustering phase transition point, the condensed phase transition point, and the interference transition point of the k -SAT problem under different values.

As a comparison of the entropy of the two-dimensional structure, Figure 9 record the changes in the number of communities that understand the space. In order to more easily observe the relationship between the changes in the number of communities in the 3-SAT problem solution space and the changes in the entropy of the two-dimensional structure, the experimental data are normalized. That is, when α is fixed, the ratio of the number of each community to the maximum number of communities is calculated, and the ratio of the two-dimensional structure entropy to the maximum two-dimensional structure entropy is also calculated in the same way. Figure 9 depict the change curve of the number of communities in the solution space of the random 3-SAT problem with $n = 100, 125, 150$, respectively. From the experimental results, it can be seen that with the increase of n , the solution space of the random 3-SAT problem becomes more complicated. The curve of the number of communities in the solution space reaches the highest point at $\alpha = 3.74$. At this time, the community structure of the solution space is at the most complex state, and the transition position of the condensed phase transition point moves from left to right. The three sets of experiments all show that the solution space community structure is the most complicated near the transition point. When $\alpha < 3.74$, the number of solution space communities increases exponentially.

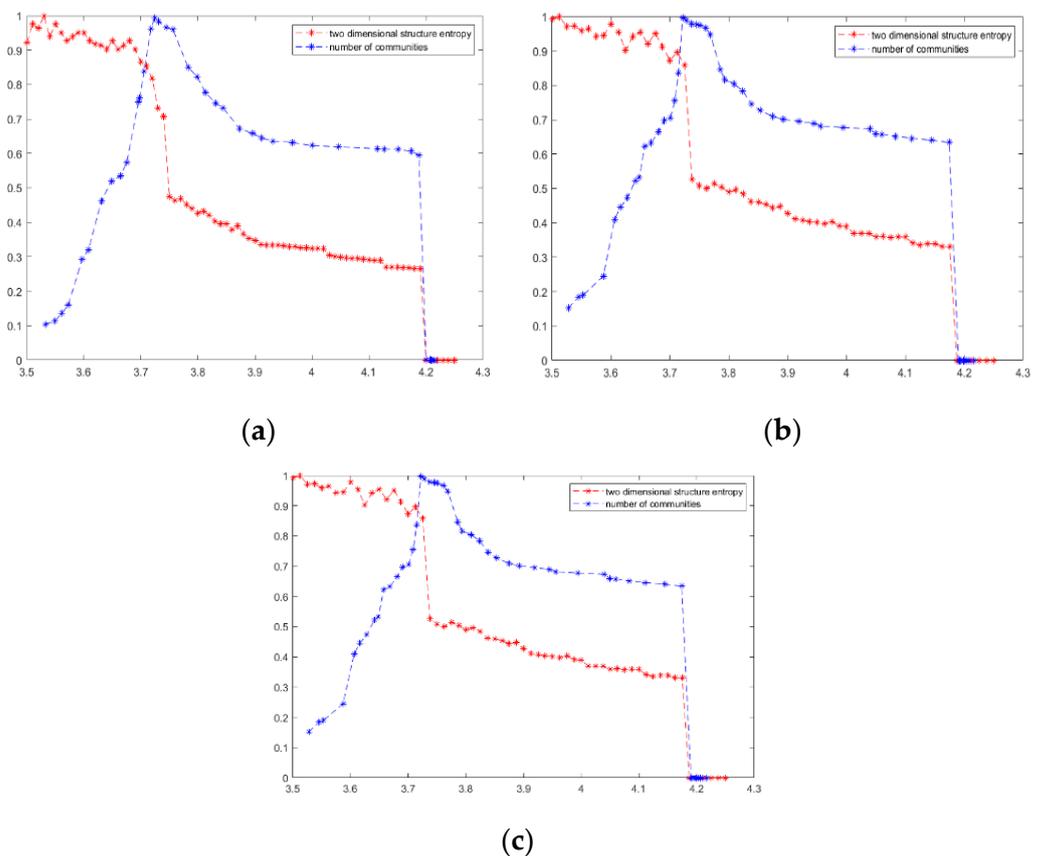


Figure 9. Changes in the Number of Communities and Two-Dimensional Structure Entropy. (a) $n = 100$ (b) $n = 125$ (c) $n = 150$.

When the solution space of SAT problem is studied by using structural entropy theory, there is no need for harsh judgment conditions, and there is no need to discuss the specific CNF in a narrow sense, only according to the optimal partition of pairs of graphs. The structural entropy theory is widely used, the judgment conditions are simple, and it can extract the multidimensional structural information of the graph model to the maximum extent, which is more effective than the methods in literature [5,6,12,13]. On the other hand, the method in this paper cannot provide a definite upper bound or lower bound, which is determined by the properties of structural entropy. Although the judgment conditions of

the method in literature [5,6,12,13] are more demanding, it can provide an exact upper and lower bound.

6. Conclusions

This paper proposes a solution space structure model based on planting strategy, using BP planting freezing variable walkSAT to structure the solution space. Considering that the original label propagation process will increase the influence of the remaining labels on the final community division result, and the original label propagation cannot divide the community well due to its strong randomness, in order to apply the results of the above construction model, LPA based on the planting strategy is proposed. The algorithm divides the community of solution spaces. Through experiments, we obtained the changes of two-dimensional structure entropy and k -dimensional structure entropy under different constraint densities of 3-SAT problems of different scales, and recorded and compared the changes of the number of communities and the two-dimensional structure entropy. The experiments showed that this solves the validity of the space analysis model. The next step will continue to focus on the solution space of the SAT problem and explore a more accurate structural entropy measurement model, so as to conduct a more detailed study on the solution space structure and multiple phase transition points of the SAT problem, provides ideas for future algorithms to solve SAT problems.

Author Contributions: Conceptualization, C.L. and X.W.; methodology, C.L. and X.W.; software, C.L., X.W. and L.L.; validation, C.L. and X.W.; formal analysis, C.L., X.W. and P.N.; investigation, C.L. and P.N.; resources, X.W.; writing—original draft preparation, C.L., X.W., L.L., P.N.; writing, review and editing, C.L., X.W., L.L. and P.N.; supervision X.W.; project administration, X.W.; funding acquisition, X.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (62062001, 61762019, 61862051, 61962002); Major Scientific Research Projects of North Minzu University (ZDZX201901); and the Natural Science Foundation of Ningxia Province of China (2020AAC03214, 2020AAC03219, 2019AAC03119, 2019AAC03120).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We would like to thank the anonymous referees for their helpful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Creignou, N.; Daudé, H. The SAT-UNSAT transition for random constraint satisfaction problems. *Discret. Math.* **2009**, *309*, 2085–2099. [[CrossRef](#)]
2. Mitchell, D.; Selman, B.; Levesque, H. Hard and easy distributions of SAT problems. *AAAI* **1992**, *92*, 459–465.
3. Martin, O.C.; Monasson, R.; Zecchina, R. Statistical mechanics methods and phase transitions in optimization problems. *Theor. Comput. Sci.* **2001**, *265*, 3–67. [[CrossRef](#)]
4. Kirkpatrick, S.; Selman, B. Critical Behavior in the Satisfiability of Random Boolean Expressions. *Science* **1994**, *264*, 1297–1301. [[CrossRef](#)]
5. Mezard, M.; Parisi, G.; Zecchina, R. Analytic and Algorithmic Solution of Random Satisfiability Problems. *Science* **2002**, *297*, 812–815. [[CrossRef](#)] [[PubMed](#)]
6. Mertens, S.; Mézard, M.; Zecchina, R. Threshold values of random K -SAT from the cavity method. *Random Struct. Algorithms* **2005**, *28*, 340–373. [[CrossRef](#)]
7. Krzakala, F.; Montanari, A.; Ricci-Tersenghi, F.; Smerjjan, G.; Zdeborova, L. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10318–10323. [[CrossRef](#)]
8. Montanari, A.; Ricci-Tersenghi, F.; Smerjjan, G. Clusters of solutions and replica symmetry breaking in random k -satisfiability. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P04004. [[CrossRef](#)]

9. Zhou, H. $T \rightarrow 0$ mean-field population dynamics approach for the random3-satisfiability problem. *Phys. Rev. E* **2008**, *77*, 066102. [[CrossRef](#)]
10. Mézard, M.; Parisi, G. The Bethe lattice spin glass revisited. *Eur. Phys. J. B-Condens. Matter Complex Syst.* **2001**, *20*, 217–233. [[CrossRef](#)]
11. Mézard, M.; Montanari, A. Reconstruction on Trees and Spin Glass Transition. *J. Stat. Phys.* **2006**, *124*, 1317–1350. [[CrossRef](#)]
12. Zhou, H. Long-Range Frustration in a Spin-Glass Model of the Vertex-Cover Problem. *Phys. Rev. Lett.* **2005**, *94*, 217203. [[CrossRef](#)] [[PubMed](#)]
13. Zhou, H. Long-range frustration in finite connectivity spin glasses: A mean-field theory and its application to the randomK-satisfiability problem. *New J. Phys.* **2005**, *7*, 123. [[CrossRef](#)]
14. Li, A.; Pan, Y. Structural Information and Dynamical Complexity of Networks. *IEEE Trans. Inf. Theory* **2016**, *62*, 3290–3339. [[CrossRef](#)]
15. Li, A.; Zhang, X.; Pan, Y. Resistance maximization principle for defending networks against virus attack. *Phys. A Stat. Mech. its Appl.* **2017**, *466*, 211–223. [[CrossRef](#)]
16. Li, A.; Yin, X.; Xu, B.; Wang, D.; Han, J.; Wei, Y.; Deng, Y.; Xiong, Y.; Zhang, Z. Decoding topologically associating domains with ultra-low resolution Hi-C data by graph structural entropy. *Nat. Commun.* **2018**, *9*, 1–12. [[CrossRef](#)] [[PubMed](#)]
17. Li, A.; Yin, X.; Pan, Y. Three-Dimensional Gene Map of Cancer Cell Types: Structural Entropy Minimisation Principle for Defining Tumour Subtypes. *Sci. Rep.* **2016**, *6*, 20412. [[CrossRef](#)]
18. Garey, M.R.; Johnson, D.S. *Computer and Intractability—A Guide to the Theory of NP-Completeness*; Freeman and Company: San Francisco, WH, USA, 1979.
19. Kschischang, F.; Frey, B.J.; Loeliger, H.A. Factor graphs and the sum-product algorithm. *IEEE Trans. Inf. Theory* **2001**, *47*, 498–519. [[CrossRef](#)]
20. Braunstein, A.; Mézard, M.; Zecchina, R. Survey propagation: An algorithm for satisfiability. *Random Struct. Algorithms* **2005**, *27*, 201–226. [[CrossRef](#)]
21. Šulc, P.; Zdeborová, L. Belief propagation for graph partitioning. *J. Phys. A Math. Theor.* **2010**, *43*, 285003.
22. Chavas, J.; Furtlehner, C.; Mezard, M.; Zecchina, R. Survey-propagation decimation through distributed local computations. *J. Stat. Mech. Theory Exp.* **2005**, *2005*, P11016. [[CrossRef](#)]
23. Li, K.; Ma, H.; Zhou, H. From one solution of a 3-satisfiability formula to a solution cluster: Frozen variables and entropy. *Phys. Rev. E* **2009**, *79*, 031102. [[CrossRef](#)] [[PubMed](#)]
24. Zhou, H. Solution space heterogeneity of the random K-satisfiability problem: Theory and simulations. *Phys. Conf. Ser.* **2010**, *233*, 012011. [[CrossRef](#)]
25. Zhou, H.; Ma, H. Communities of solutions in single solution clusters of a randomK-satisfiability formula. *Phys. Rev. E* **2009**, *80*, 066108. [[CrossRef](#)] [[PubMed](#)]