What is the structure of knowledge? Answering from the structural entropy aspect

EE447 - Mobile Internet, 2020 Spring.

∗ 姓名: 丁方玉　学号: 517030910235
∗ 姓名: 付昊源　学号: 517021910753

1 Introduction*

In this project, we applied the aspect of structural entropy to analysis the structure of knowledge. We conducted data mining on the citation network of a large number of academic papers. The description of the knowledge structure has gone through three stages from qualitative to quantitative. The most direct is to use Nebula to visualize this reference relationship network, then for the graph structure of this relationship network, we use a series of graph analysis methods to extract a tree structure of a skeleton tree from it, finally, for the obtained skeleton tree, we apply the idea of structural entropy to this skeleton tree. For each node in the tree, we can calculate the tree entropy and point entropy to reflect its influence on the entire field and its own Amount of knowledge.

2 Algorithm

2.1 Skeleton tree†

The conversion from graph to tree is actually a process of cutting edges. We use a series of graph analysis methods to cut off the extra edges, so as to retain those edges that really imply the inheritance and development of knowledge.

2.1.1 Laplacian analysis

For the method of skeleton tree extraction, it is based on graph analysis. Firstly, we can get the normalized self-loop Laplacian matrix of this graph.

$$L_{regular} = D^{-1/2}(D-W)D^{-1/2} \tag{1}$$

and then find its eigenvectors, and use the distance between the two eigenvectors to represent the distance between the two nodes.

2.1.2 Dijkstra method

After this step, our original graph structure has changed from an unweighted graph to a weighted graph. Its edges have lengths. Using Dijkstra's method, if heap optimization is used, it can be in $O(n^2 \log n)$ time. The shortest path length between each two nodes is calculated internally. Of course, there may be no path between the two nodes. In this case, we use the longest step length **maxdistance** multiplied by the average path step **averagestep** as the approximation distance between them, which is actually a very long distance.

$$d_{ij} = \begin{cases} Dijkstra(G,i,j), & \text{if there is path between i and j} \\ maxdistance * averagestep, & \text{else} \end{cases} \tag{2}$$

*Ding Fangyu, 517030910235

†Ding Fangyu, 517030910235

2.1.3 Reduction degree

After the calculation of the shortest path between nodes, we can calculate the degree of reduction of each node to the entire network. We use the concept of degree of reduction to express the relationship between two objects. If two points on an edge have different degrees of reduction to the network Large, then we can think that this edge expresses the context structure of these knowledge to a lesser extent, that is, it should be preferentially cut during the edge cutting process. Then the degree of reduction of each point to the network is actually the sum of the degree of reduction of this point relative to all other points, so how do we calculate the degree of reduction between points, using the Dijkstra result just calculated, to calculate i to the degree of reduction of j, we find the j reference article collection j_k, and sum the shortest distance of each article in i to j_k as the degree of reduction of i to j.

$$R_{ij} = \sum_{j_k} dis_{ij_k},\ RN_i = \sum_{j \neq i} R_{ij},\ \Delta RN_{ij} = |RN_i - RN_j| \tag{3}$$

2.1.4 Edge-cutting

we cut off the extra edges whose corresponding ΔRN_{ij} value is larger until the number of edges is equal to the number of nodes minus 1 through this set of processes, and what we left behind is the structure of the skeleton tree we obtained.

$$\sharp edges = \sharp nodes - 1 \tag{4}$$

2.2 Knowledge entropy*

We want to use entropy as the measure of knowledge. Entropy comes from physics, which is a thermodynamic quantity representing the unavailability of a system's thermal energy for conversion into mechanical work, often interpreted as the degree of disorder or randomness in the system. In our paper citing network, a paper's knowledge value reflects how much it influences the whole academic network. In other words, if this paper isn't published, how many other works cannot be produced. This is the target that we want to mine.

The depict of knowledge entropy starts from Shannon information entropy. This is the theoritical base of our knowledge entropy which develops the initial thinking from information entropy. From this on, we draw on the experience of structure entropy [6] [1] and propose the calculation of subtree entropy, inter-knowledge entropy and node entropy.

2.2.1 Entropy basis

The theory of entropy in information starts from Shannon entropy. Suppose there are n elements in set S, each element S_i has the probility of p_i to appear, then the entropy of S is

$$H(S) = -\sum_{i}^{n} p_i \cdot \log(p_i) \tag{5}$$

This equation is the minimum encoding length of S, it can be understood by using least knowledge value to summarize all knowledge in S. However, the shortcoming of Shannon entropy is that set is a unordered and non-structural data structure [10]. It didn't consider the inner relationship among elements, neither did it transfer to graph structure.

*Fu Haoyuan, 517021910753

Based on these shortcomings, *Li et al* [6] proposed structure entropy to migrate Shannon entropy on graphs. Starting from a skeleton tree, structure entripy $H(G)$ is defined

$$H(G) = -\sum_{\alpha \in T} \frac{g_\alpha}{2m} \log \frac{V_\alpha}{V_{\alpha^-}} \tag{6}$$

where T is the skeleton tree extracted from arbitrary graph G, g_α is the cut set of all nodes in sub-tree whose root is α, $m = ||V(G)||$, V_α and V_{α^-} is the number of nodes in sub-tree whose root is α and parent node of α, respectively.

Note that the structure entropy is defined on skeleton tree, rather than graph, this is the reason why we need to propose the algorithm of extracting skeleton tree from initial paper citing graph.

2.2.2 Sub-tree entropy

Sub-tree entropy is designed to measure the influence of a paper to the whole academic network, in our project, to the whole Cora dataset. Thus, a paper's sub-tree entropy should consider the sub-tree whose root is the paper. The algorithm of sub-tree entropy is based on structure entropy. Since we want to measure the whole subtree, we preserve the initial form of structure entropy and define sub-tree entropy as

$$H^T(\alpha) = -\frac{g_\alpha}{2m} \log \frac{V_\alpha}{V_{\alpha^-}} \tag{7}$$

The notations keep same as equation 6, and we omit here. We can further consider sub-tree entropy as a paper's uncertainty to whole academic dataset. A bigger sub-tree entropy represents a bigger influence to this academic field. We also emphasize a misunderstanding here: the definition and physical meaning of sub-tree entropy may fuse people that the number of nodes in a sub-tree has positive correlation with sub-root entropy. In fact, the most import parameter is the cut set of g_α, but not V_α. The following experiments will also show that some papers with small number of nodes in sub-tree also have rather high sub-tree entropy.

2.2.3 Inter-knowledge entropy

The inter-knowledge entropy is defined in the range of papers with same parent node in skeleton tree. Since they are from the same parent node, they should have some knowledge in common, and here we want to extract this value. The definition is

$$I(a,b) = -\frac{g_{ab}}{4m} \log \frac{V_a V_b}{V_{ab^-}^2} \tag{8}$$

where g_{ab} is the cut set of $T(a) \vee T(b)$, and V_{ab^-} is the common parent node of paper a and b.

Our proposed inter-knowledge entropy satisfies two intuitively correct properties:

- Symmetry: $I(a,b) = I(b,a)$
- Self-symmetry: $I(a,a) = H^T(a)$, where $H^T(a)$ is the sub-tree entropy of paper a.

The proof of them is easy and I omit it here. In physical meaning, inter-knowledge entropy depicts the correlation between two papers a and b, since this correlation is defined by their jointly parent, we only define this entropy among child nodes of an arbitrary node. Nodes with different direct parent node do not have this inter-knowledge entropy. In real calculation, this definition is enough since we have already considered the correlation among nodes with different parent nodes in sub-tree entropy. And we'll show in node entropy that the inter-knowledge entropy is just an auxiliary value.

2.2.4 Node entropy

Based on sub-tree entropy and inter-knowledge entropy, we can define node entropy, which is the measure of knowledge value of any given node a. This calculation starts from a's sub-tree entropy $H^T(a)$, and we delete all sub-tree entropy of a's child nodes, because these entropy doesn't directly belong to a. Besides, purely deleting these entropy seems rigid and doesn't consider the correlation among these papers. To fix this problem, we add the sum of inter-knowledge entropy among all nodes whose parent is a. This is because we regard the correlation among children papers defined by the parent paper. Thus, in formulation, the calculation of node entropy is

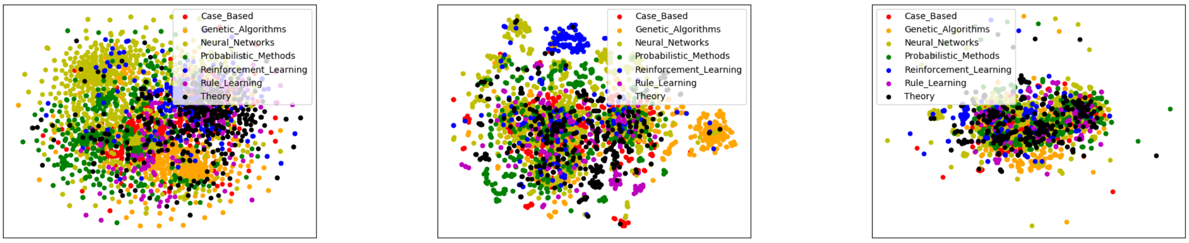
$$H^N(a) = |H^T(a) - \sum_{a_i \in T(a)} H^T(a_i) + \sum_{a_i \in T(a)} \sum_{a_j \in T(a), j \neq i} I(a_i, a_j)| \tag{9}$$

In conclusion, we compare the relationship and difference among three entropies above. Sub-tree entropy imitate the calculation of structure entropy, depicting how much influence a paper to the whole academic field. Inter-knowledge is defined among nodes with same direct parent node, measuring the correlation of a paper pair. Node entropy considers sub-tree entropy and inter-knowledge entropy together to calculate the real value of knowledge for a given paper.

3 Experiments

3.1 Initial clustering in Cora[*]

To measure the knowledge embedded in a graph, we first tried to parse the clustering or classification features. We think a good measurement should at least correctly and clearly classify papers from different academic fields in cora [8]. Under this motivation, we first extract three naive features from initial graph: keyword feature, citing feature and spectral clustering feature. Keyword features come from cora dataset itself, using 0/1 to represent a keyword's appearance and packaged to a vector. Citing feature is the citing relationship of papers in cora dataset, using 0/1 to represent whether this paper cites or is cited. Spectral clustering feature is initially from citing feature, but using spectral clustering algorithm [9] to get a Laplace matrix, and we treat each row as feature. To visualize these features and check their clustering effect, we use t-SNE as a visualization in scatter graph. The results are shown in 1. Obviously these naive features are not what we want, and we need more robust methods to mine knowledge embedded in graphs deeply.



(a) Keyword feature visualization.

(b) Citing feature visualization.

(c) Spectral clustering feature visualization.

Fig. 1: Different naive features visualized using t-sne algorithm. Features are messy and many papers in different academic fields mix together. It seems we must use more powerful and complex algorithms to further extract and analysis these features.

[*] Fu Haoyuan, 517021910753

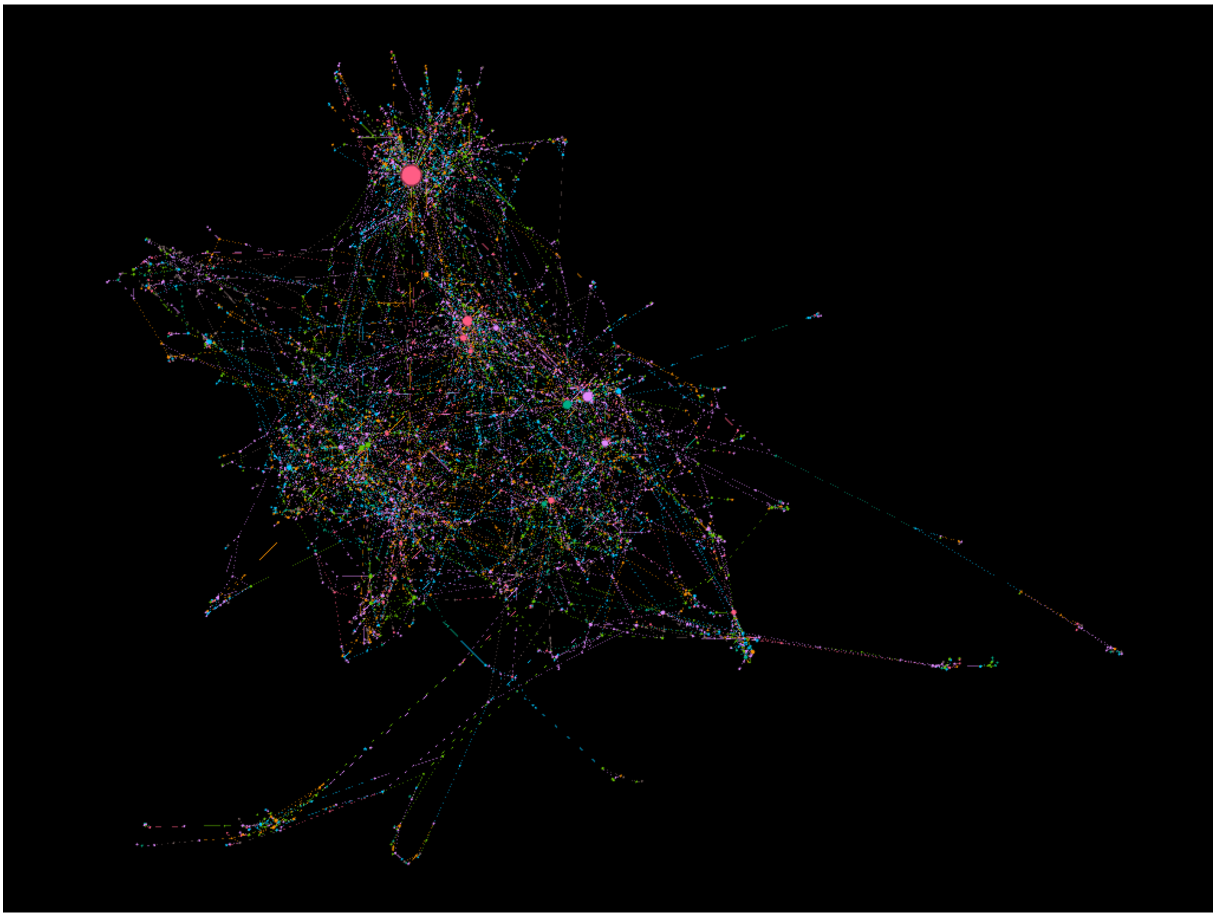


Fig. 2: Cora Nebula.

3.2 Nebula of Cora[*]

The academic papers we visualized with gephi cited the nebula of the network. It is actually a very qualitative depiction of the knowledge structure. The nebula drawn is largely dependent on the visualization method of gephi. On the whole, it shows a very Approximate result.

3.3 Skeleton tree extraction[†]

After we cut off the extra edges through the skeleton tree extraction processes, what we left behind is the structure of the skeleton tree we obtained. Compared to the original nebula map, the description on the structure of knowledge and the development of knowledge are much clearer.

3.4 Sub-tree entropy based tree[‡]

From the skeleton tree extracting algorithm, we have already extracted a skeleton tree from initial cora dataset graph. However, all nodes in tree are equivalent and we cannot focus on the stress. To stress the main points, we use adjust the radius of each node according to its sub-tree entropy. A node with higher sub-tree entropy will have larger radius. Since the number of nodes is large, and an overall image will be nasty and misunderstanding, we only capture some screenshots here to emphasize the nodes with higher sub-tree entropy. The visualization result is like figure

[*] Ding Fangyu, 517030910235

[†] Ding Fangyu, 517030910235

[‡] Fu Haoyuan, 517021910753

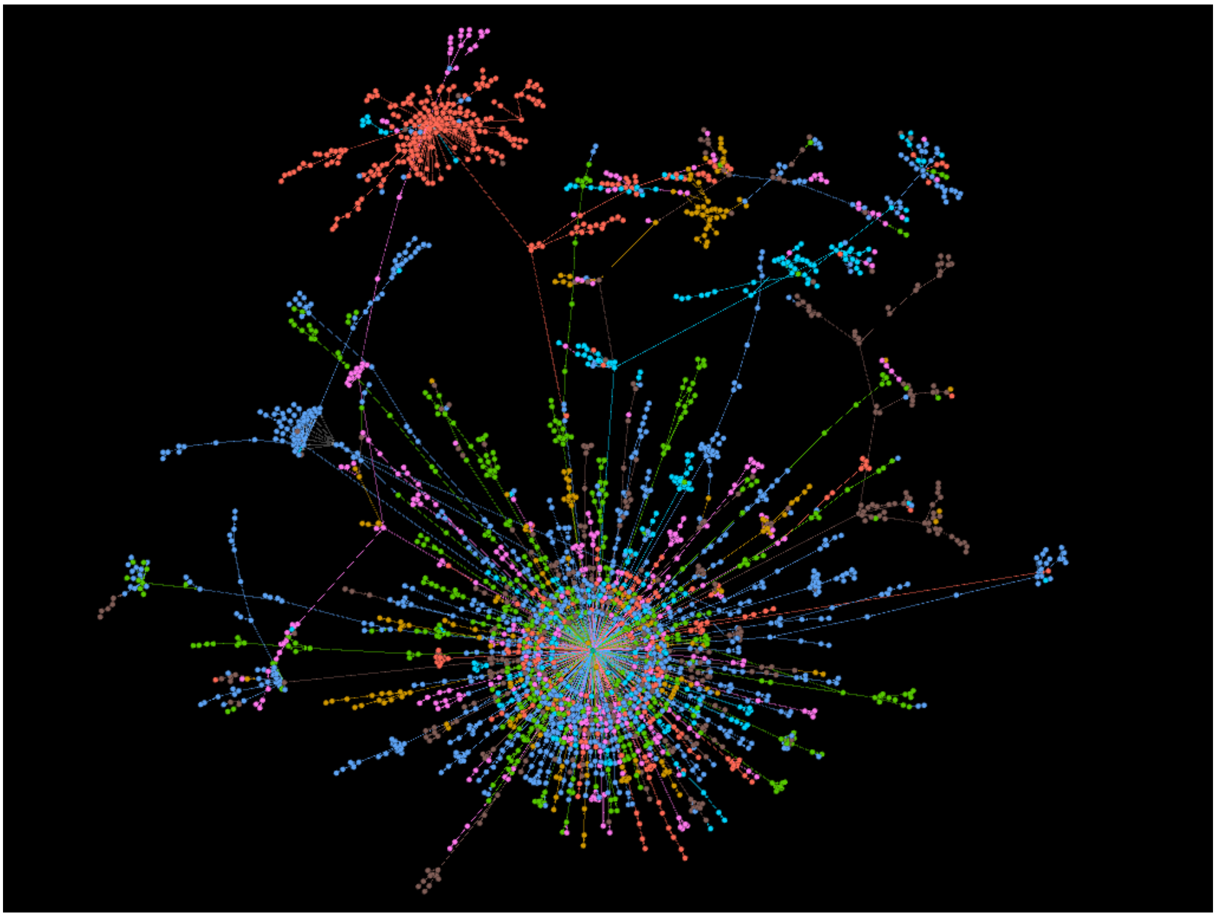
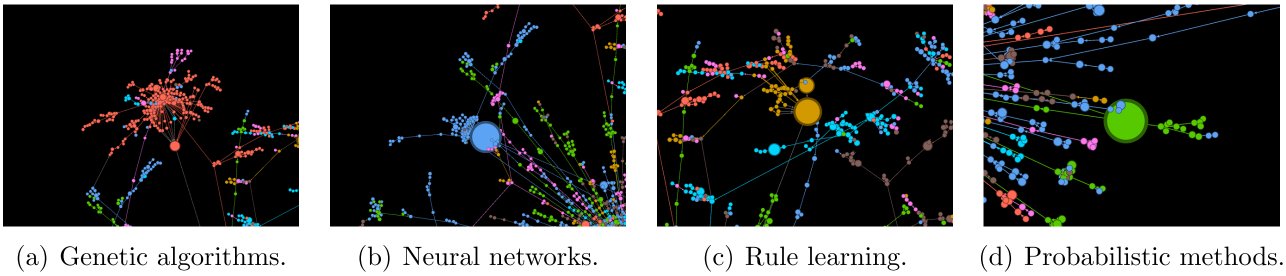


Fig. 3: Skeleton tree, where different colors replace different fields, the clustering property is clearly shown.



(a) Genetic algorithms.

(b) Neural networks.

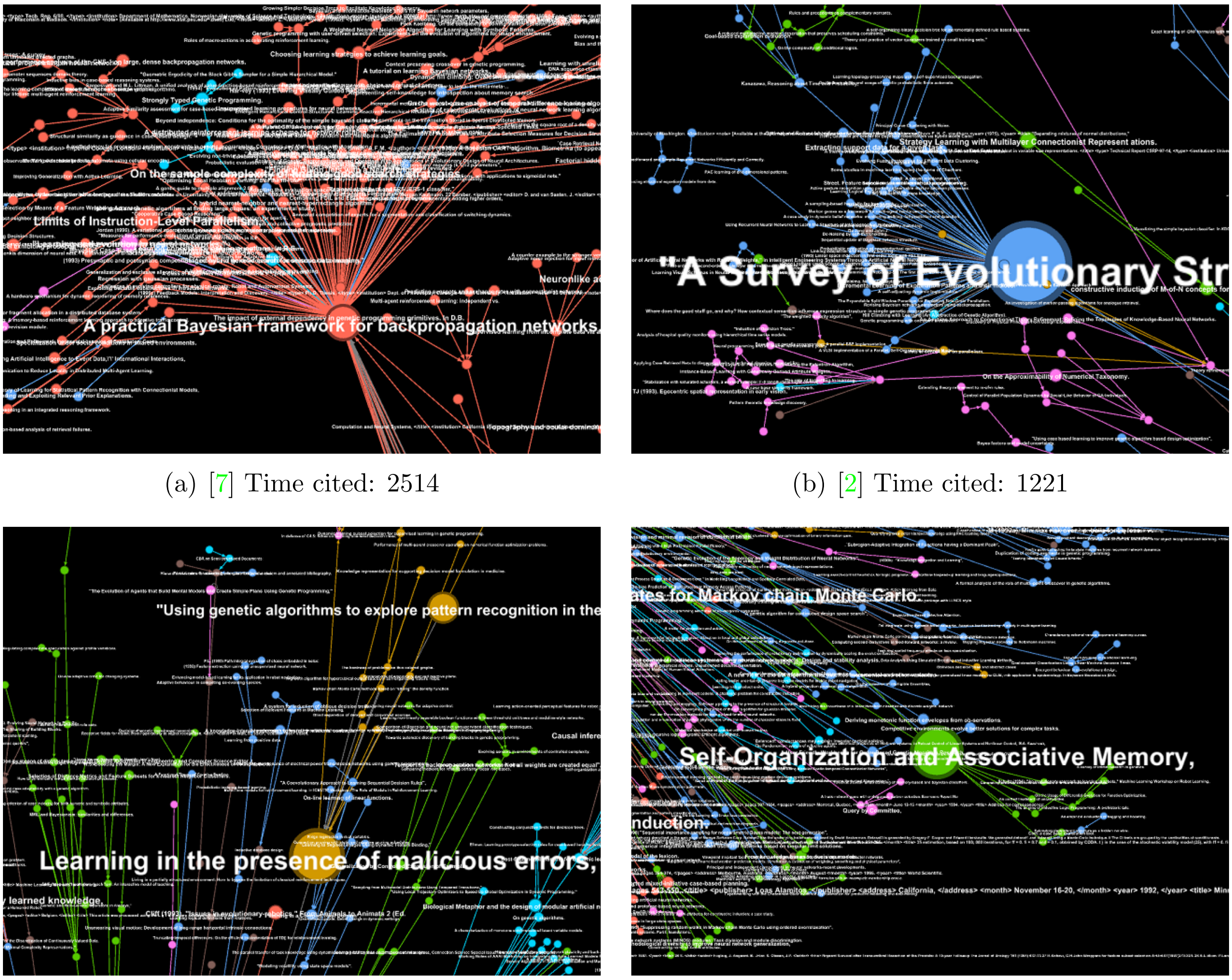
(c) Rule learning.

(d) Probabilistic methods.

Fig. 4: Local screenshots of skeleton tree, with each node's radius corresponding to its sub-tree entropy. The color of the emphasized node means different academic fields, which are shown in the sub-title of each image.

4. The result show that a big node will usually have a cluster of nodes in same academic field surrounding it. The cluster is the sub-tree of the node with high sub-tree entropy.

To quantitatively analyse the correctness our sub-tree entropy algorithm, we labeled each node with its title, whose font size also has positive corresponding to its sub-tree entropy. Thus, we can search the real citing times in google, and check whether these papers with high sub-tree entropy are cited for rather many times. The visualization and citing times result are shown in figure 5. Note that the structure of tree is different from figure figure 4, because we used a new layout in gephi [3] software to thin out nodes so that we can clearly see each title.


A practical Bayesian framework for backpropagation networks.
Limits of Instruction-Level Parallelism.
On the sample complexity
"A Survey of Evolutionary Str
"Using genetic algorithms to explore pattern recognition in the
Learning in the presence of malicious errors,
ates for Markov chain Monte Carlo.
Self-Organization and Associative Memory,
nduction.

(a) [7] Time cited: 2514

(b) [2] Time cited: 1221

(c) [4] Time cited: 519

(d) [5] Time cited: 16055

Fig. 5: Local screenshots with title labeled on the nodes. We have also searched on google scholar for the time cited of each paper with extremely high sub-tree entropy. All emphasized nodes in images above are with rather high time cited.

3.5 Relationship of knowledge entropy[§]

In this experiment, we want to check the relationship between sub-tree entropy and node entropy of a given node. From previous analysis, sub-tree entropy and node entropy depict knowledge entropy from two different perspective. Intuitively, a paper that produces huge effect to the whole academic field should also have informative knowledge itself. To check this point, we use scatter graph to study the relationship between these two kinds of entropy. In our experiment, we found

[§] Fu Haoyuan, 517021910753

that the value of entropy has big range from 10^{-7} to 10^3, which is unreadable in a normal coordinate. To normalize this range, we use log-log scale coordinates and project each node in it. The result in figure 6 shows the approximatedly positive correlation between sub-tree entropy and node entropy.

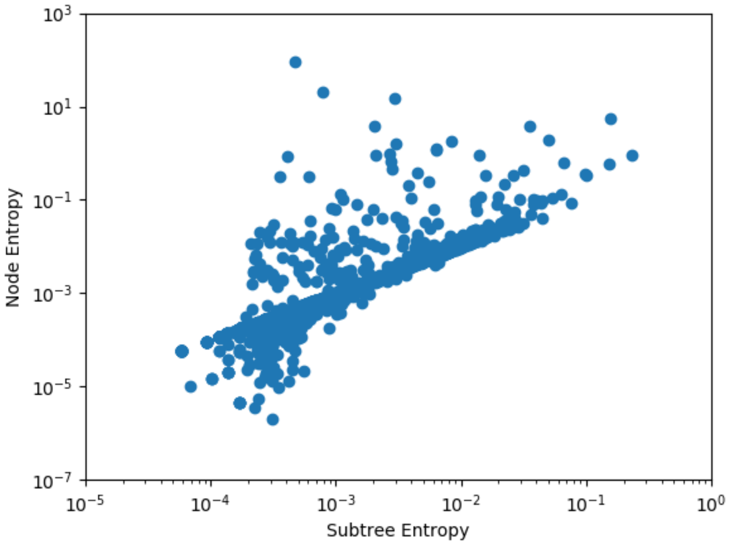


Fig. 6: The scatter graph of each paper with corresponding node entropy and sub-tree entropy. The tendency of this graph shows the positive relation between these two kinds of entropy.

3.6 Relationship of level and entropy[¶]

In this experiment, we want to study the relationship between a node's level in skeleton tree and its knowledge entropy, including sub-tree entropy and node entropy. We first calculated the average sub-tree entropy and node entropy in each level of skeleton tree, then use a line chart to show the result in figure 7.

However, as we have mentioned before, value of entropy varies a lot, and an outlier point may unpredictably influence the average value. This is the reason why line chart in figure 7 fluctuates terribly. For more precise results, we still use log-scale coordinate with scatter graph, and draw each data point on the figure. Figure 8 shows this result and make some illustration.

4 Conclusion and future work[‖]

In this project, we depict knowledge from the perspective of graph structure. We take cora dataset as the initial academic network graph, then proposed a skeleton extracting algorithm based on spectral clustering. After extracting the skeleton tree, we further calculated the sub-tree entropy, inter-knowledge entropy and node entropy of each paper in cora dataset. Finally, we did a series of experiments on this graph to verify our hypothesis and dig some relationship between values in our work. The experiment results show that our algorithm makes sence and satisfy our intuition in many aspects.

However, this work is not perfect. Our algorithm can only work on a given graph, and no changes can be considered as time goes on. If you want an online version algorithm, you can only

[¶] Fu Haoyuan, 517021910753

[‖] Fu Haoyuan, 517021910753

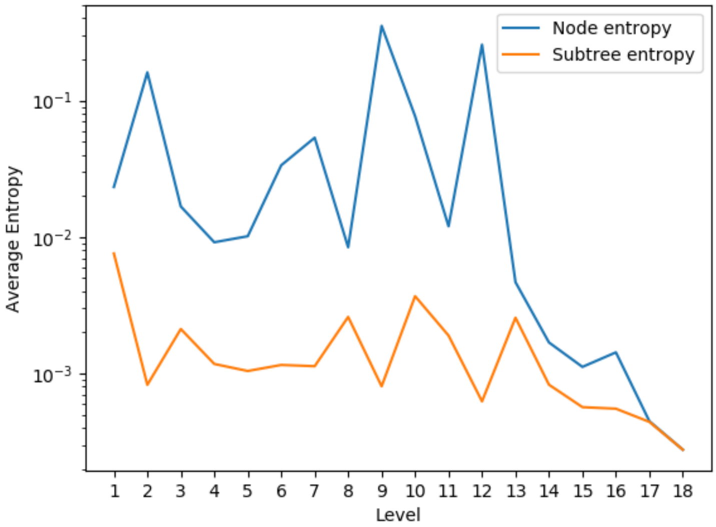
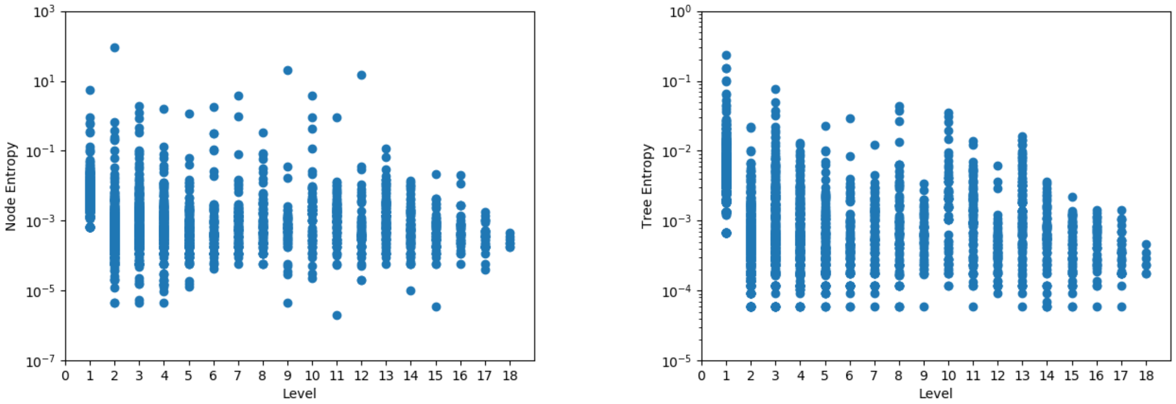


Fig. 7: Relationship between average knowledge entropy and level. The average entropy is calculated in each level and no weight.



(a) Node entropy - level scatter graph.

(b) Sub-tree entropy - level scatter graph.

Fig. 8: These two figures show the scatter graph with relationship between entropy and level in skeleton tree. The upper bound of entropy has the tendency to fall as the level increases. In other words, the level in skeleton tree determines the superior limit of entropy. If a paper wants to produce more influence to the whold academic field, it must be near the initial paper of this field in skeleton tree.

redo the whold story again. Thus, it misses the function of predicting a paper's development. Besides, our CPU-based method cannot deal with graphs having more than 10,000 nodes due to limit of memory. Moreover, we didn't consider the content of a paper, which is also an essential index in evaluating knowledge.

For future work, we want our algorithm to be migrated to other networks, including social network, market network and rumor network, etc. Besides, since the data structure of graph is the most common in human real life, we want this algorithm to be more systematical and general, but not designed only for this task.

Reference

[1] Kartik Anand and Ginestra Bianconi. Entropy measures for networks: Toward an information theory of complex topologies. *Physical Review E*, 80(4):045102, 2009.

[2] Thomas Back, Frank Hoffmeister, and Hans-Paul Schwefel. A survey of evolution strategies. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 2–9. Morgan Kaufmann, 1991.

[3] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. In *Third international AAAI conference on weblogs and social media*, 2009.

[4] Michael Kearns and Ming Li. Learning in the presence of malicious errors. *SIAM Journal on Computing*, 22(4):807–837, 1993.

[5] Teuvo Kohonen. *Self-organization and associative memory*, volume 8. Springer Science & Business Media, 2012.

[6] Angsheng Li and Yicheng Pan. Structure entropy and resistor graphs. *arXiv preprint arXiv:1801.03404*, 2018.

[7] David JC MacKay. A practical bayesian framework for backpropagation networks. *Neural computation*, 4(3):448–472, 1992.

[8] Andrew McCallum. Cora dataset. 2017.

[9] Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

[10] Mirjam Weilenmann and Roger Colbeck. Analysing causal structures with entropy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2207):20170483, 2017.

Table. 1: Division of work

Model & report	
Skeleton tree part	Ding Fangyu
Structural entropy part	Fu Haoyuan