ICIC Express Letters
Volume **11**, Number **4**, April **2017**

ICIC International ©2017 ISSN 1881-803X
pp. **823–828**

TWO THEOREMS ON NODES OF CONCEPT LATTICE AND PARTIAL-ORDER FORMAL STRUCTURE

Wenxue Hong[1,*], Yunli Ren[1,2], Cunfang Zheng[1], Jialin Song[1] and Shaoxiong Li[1]

[1]School of Electrical Engineering
Yanshan University
No. 438, Hebei Avenue, Qinhuangdao 066004, P. R. China
*Corresponding author: hongwx@ysu.edu.cn

[2]College of Mathematics and Information Technology
Hebei Normal University of Science and Technology
No. 360, Hebei Avenue, West Section, Qinhuangdao 066004, P. R. China

Received September 2016; accepted December 2016

Abstract. *Formal concept analysis, as an important tool for data analysis and knowledge processing, has been applied to many fields. However, the visual effect of its Hasse diagram is not satisfactory especially for a large or complex formal context, since edges may intersect each other. To solve this problem, partial-order formal structure (PFS for brief) is proposed and has been successfully applied to the field of data analysis. This paper studies mainly on the nodes of PFS from the perspective of comparing with concept lattice. Specifically, a necessary and sufficient condition of a node of PFS to be a pseudo-concept, and two sufficient conditions on which the node set of PFS is a proper subset of that of concept lattice are proposed and proved in detail.*
Keywords: Formal concept analysis, Concept lattice, Partial-order formal structure, Node, Pseudo-concept

1. **Introduction.** Formal concept analysis (briefly FCA) was initially proposed by Wille in 1982 [1]. And in 1999, Ganter and Wille summarized the early framework and theoretical results of FCA in their works *Formal Concept Analysis: Mathematical Foundations* [2]. After decades of developments, FCA now has evolved into an important branch of applied mathematics and been widely used in various areas, such as knowledge discovery, data analysis, text retrieval and software engineering [3-8].

There are two key components in FCA: formal concept and concept lattice. For a formal context K, a formal concept is a pair of an object subset (extension) and an attribute subset (intension), which determine each other mutually. A partial order among the formal concepts can be established according to the inclusion relationship of their extensions. All the formal concepts together with their partial order form a complete lattice called concept lattice, which shows the hierarchical organization of concepts and is the core of FCA. Hasse diagram is introduced to visualize the hierarchical relation among formal concepts. However, the visual effect of Hasse diagram is not satisfactory especially for a large or complex formal context, since a lot of edges may intersect each other.

Concerning this problem, our team put forward the idea of partial-order formal structure (PFS) [9]. PFS, with nodes similar to those of Hasse diagram of concept lattice, is a closed, acyclic tree diagram. Thus, there is no cross of lines and the visual effect is perfect. Meanwhile, PFS is a good tool of knowledge discovery and rules extraction since it can illustrate the universality and specificity of attributes (objects). And scholars have successfully applied it in knowledge discovery of Chinese traditional medicine and languages [10-12].

When comparing attributes of objects, we find that some are universal while some are more specific. It is the essence of cognizing and distinguishing objects to find their specificities out of universial common attributes. PFS, based on the fundamental principle of human's cognition, is just such a new theory that takes exploration of attributes' relationships and distinction of objects as its basic purpose, while FCA is a strict mathematical theory that studies on the relationships among concepts. Although the two theories have different focuses, they are both powerful tools in data mining, similar to each other but possessing their own advantages. Thus, in order to make better use of them, it is necessary to give a comparison between concept lattice and PFS.

In this paper, we study mainly on the nodes of PFS, from the perspective of comparing them with those of concept lattice. And the paper is organized as follows. In Section 2, formal contexts and concept lattice are reviewed. Section 3 introduces the constructing method of PFS. And Section 4 focuses on the nodes of PFS. Specifically, a necessary and sufficient condition of a node of a PFS to be a pseudo-concept, and two sufficient conditions on which the node set of PFS is a proper subset of that of concept lattice are proposed and proved in detail. Finally, main results of the paper are summarized in Section 5.

2. **Formal Contexts and Concept Lattice.** In this section, we review the basic definitions regarding concept lattice.

Definition 2.1. [13] *A formal context $K = (G, M, I)$ consists of two sets G and M and a relation I between them. The elements of G and M are called objects and attributes respectively. $(g, m) \in I$ or gIm denotes that the object g possesses the attribute m.*

For subsets $A \subseteq G$ and $B \subseteq M$, define

$$f(A) = \{m \in M | xIm \; for \; all \; x \in A\}, \; g(B) = \{x \in G | xIm \; for \; all \; m \in B\}.$$

Definition 2.2. [13] *A formal concept of a formal context $K = (G, M, I)$ is a two-tuple (A, B), where $A \subseteq G$, $B \subseteq M$, and $f(A) = B$, $g(B) = A$. A and B are the extension and intension of the concept (A, B) respectively.*

Definition 2.3. [13] *If (A_1, B_1) and (A_2, B_2) are two concepts of a formal context, and $A_1 \subseteq A_2$ (i.e., $B_2 \subseteq B_1$), then (A_1, B_1) is called the sub-concept of (A_2, B_2), and (A_2, B_2) is called the sup-concept of (A_1, B_1), denoted by $(A_1, B_1) \leq (A_2, B_2)$. All the formal concepts of the formal context $K = (G, M, I)$ with the above order form a complete lattice called concept lattice, denoted by $B(G, M, I)$.*

3. **Constructing Method of Partial-Order Formal Structure.** If an object x satisfies that $f(x) = M$, there is little significance to consider it. Therefore, we omit such objects from formal contexts.

The main procedures of constructing PFS of a formal context are as follows.

(1) The maximal common attribute m is found, i.e., the attribute m satisfying that $g(m) = G$. Accordingly, the top node $(m, g(m))$ is acquired (Here m is used to replace the attribute subset $\{m\}$ for simplity. Similar processings will not be mentioned hereinafter.). If there is no such an attribute, then the top node is denoted by (Φ, G).

(2) The attributes of the simplest cover of G are found, i.e., attributes $m_1, m_2, \ldots, m_k$ satisfying the following items:

i) $\bigcup_{i=1}^{k} g(m_i) = G$;

ii) For every $l \in \{1, 2, \ldots, k\}$, $\bigcup_{i=1, i \neq l}^{k} g(m_i) \neq G$;

iii) $\sum_{i=1}^{k} |g(m_i)|$ is the largest among attributes satisfying item i) and item ii);

iv) k is the minimal positive integer satisfying item i) and item iii).

(3) The first-layer nodes $(m_i, g(m_i))$, $i = 1, 2, \ldots, k$ are established, and located under the top node. Lines are drawn between the top node and every first-layer node. If there exists the maximal common attribute m, then m should be added to these nodes, i.e., $(mm_i, g(mm_i))$, $i = 1, 2, \ldots, k$.

(4) For every m_i $(i = 1, 2, \ldots, k)$, the corresponding sub-context $K_{m_i} = (g(m_i), M, I_{m_i})$ can be got, where I_{m_i} denotes the relation $I \cap (g(m_i) \times M)$. Under the sub-contexts K_{m_i}, the attributes of the simplest cover of $g(m_i)$ $(i = 1, 2, \ldots, k)$ are found sequentially, and denoted by $m_1^i, m_2^i, \ldots, m_{s(i)}^i$ $(s(i) \in N^+,\ i = 1, 2, \ldots, k)$. And then the corresponding nodes $\left(m_i m_j^i, g\left(m_i m_j^i\right)\right)$ are acquired $(j = 1, 2, \ldots, s(i),\ i = 1, 2, \ldots, k)$. The above nodes constitute the second-layer nodes, located under the first-layer ones. Lines are drawn between them and their corresponding first-layer nodes.

Note: i) If the objects of a node are included in the object sets of some already-existing nodes, then the node is omitted.

ii) If $\bigcup\limits_{i=1}^{k} g(m_i) \neq \bigcup\limits_{i=1}^{k} \bigcup\limits_{j=1}^{s(i)} g\left(m_i m_j^i\right)$, then t lines are drawn between the first-layer nodes and the bottom node (M, Φ), where $t = \left|\bigcup\limits_{i=1}^{k} g(m_i) - \bigcup\limits_{i=1}^{k} \bigcup\limits_{j=1}^{s(i)} g\left(m_i m_j^i\right)\right|$.

(5) Step (4) is repeated until no new node can be formed under the existing nodes. And then lines are drawn between the nodes and the bottom node (M, Φ).

Example 3.1. *The following Table 1 is the formal context of lives and water which is typical in [13]. The universe is* $G = \{$*fishleech, bream, frog, dog, waterweeds, reed, bean, corn*$\}$*, elements of which are denoted by numbers from 1 to 7, and the attribute set is* $M = \{$*needs water to live, lives in water, lives on land, needs chlorophyll, dicotyledon, monocotyledon, can move, has limbs, breast feeds*$\}$*, elements of which are denoted by letters* a, b, c, d, e, f, g, h, i, j *respectively. If an object possesses an attribute, then a cross is drawn in the intersection of the row where the object is located and the column where the attribute is located. The corresponding concept lattice and PFS are shown in Figure 1 and Figure 2 respectively.*

Table 1. Formal context of lives and water

	a	b	c	d	e	f	g	h	i
1 fishleech	×	×					×		
2 bream	×	×					×	×	
3 frog	×	×	×				×	×	
4 dog	×		×				×	×	×
5 water weeds	×	×		×		×			
6 reed	×	×	×	×		×			
7 bean	×		×	×	×				
8 corn	×		×	×		×			

From Figure 1 we see that some edges of the concept lattice intersect with each other. In Figure 2, PFS is a closed, acyclic tree diagram without edges' intersection and is simpler.

Although there are less edges and nodes in Figure 2, the hierarchy of attributes is illustrated clearly. Specifically speaking, the attribute a*, possessed by every object, is the most common one and therefore is located on the top. Comparing among other attributes except* a*, the attributes* b *and* c *are possessed by more objects and thus placed on the first layer under* a*. Similarly the rest attributes are located in the same way.*

As for the nodes of the two figures, except for (acg, 34) and (abd, 56), nodes in Figure 2 are included in Figure 1, while nodes (ag, 1234), (ad, 5678), (agh, 234), (adf, 568)

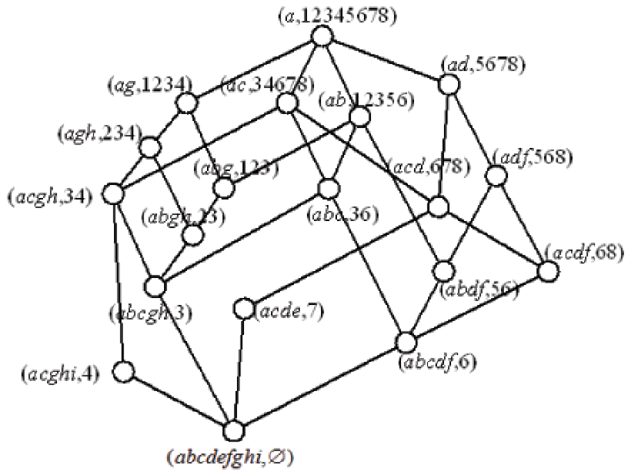

FIGURE 1. Concept lattice

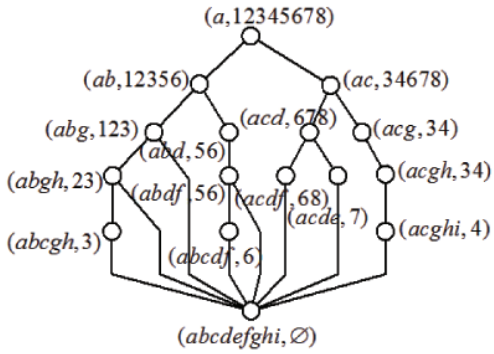

FIGURE 2. PFS

and (abc, 36) of Figure 1 do not appear in Figure 2. The relationship between nodes of concept lattice and PFS which seems complicated will be discussed in the next section.

4. **Comparison of Nodes between PFS and Concept Lattice.** From Section 3, we know that all the nodes of PFS have such a form as $(B, g(B))$, where $B \subseteq M$. Generally speaking, it is not definitely a formal concept, for the equality $f(g(B)) = B$ may not hold sometimes. Then we are concerned about how to determine from the diagram of PFS whether a node is a concept or not. And the following Theorem 4.1 and Corollary 4.1 will solve it.

Theorem 4.1. *In PFS, the node $(B, g(B))$ is a pseudo-concept if and only if there is only one branch stemming from $(B, g(B))$ and the object set of the node under it is also $g(B)$.*

Proof: Firstly, the necessity is to be proved. Suppose that $(B, g(B))$ is a pseudo-concept. Then $f(g(B)) \supset B$, i.e., there is an attribute $a \in M$ such that $a \in f(g(B))\backslash B$. Since $a \in f(g(B))$, we have that $g(a) \supseteq g(f(g(B))) = g(B)$. Thus, $g(a \cup B) = g(a) \cap g(B) = g(B)$. According to the constructing rule of PFS, we know that $g(a \cup B)$ is the simplest cover of $g(B)$ and then the node $(a \cup B, g(a \cup B)) = (a \cup B, g(B))$ is the unique one under $(B, g(B))$. Therefore, there is only one branch stemming from $(B, g(B))$ and the object set of the node $(a \cup B, g(B))$ under it is also $g(B)$.

Next the sufficiency is to be proved. Suppose that there is only one branch stemming from $(B, g(B))$ and the object set of the node under it is also $g(B)$. Then it is reasonable to suppose that the node under $(B, g(B))$ is $(a \cup B, g(B))$. Accordingly, $f(g(B)) = f(g(a \cup B)) \supseteq a \cup B \supset B$. Therefore, $f(g(B)) \neq B$, which means that $(B, g(B))$ is a pseudo-concept.

Corollary 4.1. *In PFS, the node $(B, g(B))$ is a formal concept if and only if there are at least two branches stemming from it or there is only one branch stemming from $(B, g(B))$ but the object set of the node under it is properly included in $g(B)$.*

Proof: It is straightforward according to Theorem 4.1.

From Corollary 4.1, we know that the node set of PFS is a subset of that of concept lattice, if we take no consideration of pseudo-concepts. Now there is an interesting question: provided that pseudo concepts are not considered, when is the node set of PFS a proper subset of that of concept lattice and when are they the same? Before discussing this question, the definition of multi-attribute associated attribute is to be introduced.

Definition 4.1. *In a formal context $K = (G, M, I)$, the attribute m_1 is called a multi-attribute associated attribute of $m_2, m_3, \ldots, m_k$, if the attributes $m_1, m_2, \ldots, m_k$ satisfy the following items:*

(1) Any two elements of $\{g(m_1), g(m_2), \ldots, g(m_k)\}$ do not have the inclusion relation;

(2) $g(m_1) \subseteq \bigcup_{i=2}^{k} g(m_i)$.

Theorem 4.2. *In the PFS of a formal context $K = (G, M, I)$, if for the attributes of the simplest cover of G, there is a multi-attribute associated attribute, then the node set of PFS is a proper subset of that of concept lattice, without considerations of pseudo-concepts.*

Proof: Suppose that for the attributes $m_1, m_2, \ldots, m_k \in M$, their object sets $g(m_1)$, $g(m_2), \ldots, g(m_k)$ form the simplest cover of G. Then $(m_1, g(m_1))$, $(m_2, g(m_2)), \ldots$, $(m_k, g(m_k))$ constitute the first-layer nodes of PFS. Suppose that $m \in M$ is a multi-attribute associated attribute of $m_1, m_2, \ldots, m_k$. Then

$$g(m) \subseteq \bigcup_{i=1}^{k} g(m_i), \tag{1}$$

and for arbitrary $i \in \{1, 2, \ldots, k\}$,

$$g(m) \not\subset g(m_i). \tag{2}$$

It follows from (1) that in PFS, m is located under the attributes $m_1, m_2, \ldots, m_k$. That is to say, in PFS, nodes containing m only appear in such a form as $(B \cup m_i \cup m, g(B \cup m_i \cup m))$, where $B \subseteq M$ may be null and $m_i \in \{m_1, m_2, \ldots, m_k\}$.

On the other hand, from (2) it can be inferred that the concept $(f(g(m)), g(m))$ is in the concept lattice but not in PFS (Here the positions of the object set and the attribute set are exchanged for consistency with PFS.). It can be proved by contradiction. Suppose that $(f(g(m)), g(m))$ is a node in PFS. Then according to the above statements, there is some $m_i \in \{m_1, m_2, \ldots, m_k\}$, such that $(f(g(m)), g(m)) = (B \cup m_i \cup m, g(B \cup m_i \cup m))$. Hence $m_i \in f(g(m))$, which means that $\{m, m_i\} \subseteq f(g(m))$. Therefore,

$$g(m) = g(f(g(m))) \subseteq g(m, m_i) = g(m) \cap g(m_i) \subseteq g(m_i).$$

It follows that $g(m) \subseteq g(m_i)$, which contradicts with (2). So the concept $(f(g(m)), g(m))$ is in the concept lattice but not in PFS, and thus Theorem 4.2 holds.

Corollary 4.2. *Suppose that $(B, g(B))$ is a node in the PFS of a formal context $K = (G, M, I)$, and it is the first one among the nodes in the same layer. If for the attributes of the simplest cover (or sub-universe cover) of $g(B)$, there is a multi-attribute associated attribute in the sub-context $(g(B), M, I \cap g(B) \times M)$, then the node set of PFS is a proper subset of that of concept lattice, without considerations of pseudo-concepts.*

The proof of Corollary 4.2 is similar to that of Theorem 4.2, so it is omitted.

Now let us return to Example 3.1. From Table 1, we know that $g(b)$ and $g(c)$ form the simplest cover of G, and attributes d and g are multi-attribute associated attributes of b

and c. Then according to the proof of Theorem 4.2, in PFS, nodes containing d or g do not appear unless b or c is included in the attribute set. From Figure 1 and Figure 2, we get that the formal concepts (ag, 1234) and (ad, 5678) do not appear in the PFS (Figure 2) indeed, which coincide with Theorem 4.2.

5. **Conclusions.** This paper focuses mainly on the nodes of PFS from the perspective of comparing with concept lattice. Firstly, a necessary and sufficient condition of a node of PFS to be a pseudo-concept is proved, which indicates that in PFS, every node in PFS is a formal concept, except for the nodes, each of which has only one branch stemming from it and whose object set is the same as that of the node under it. This result may provide a new method to find formal concepts from a formal context. Secondly, by means of multi-attribute associated attributes, two sufficient conditions on which the node set of PFS is a proper subset of that of concept lattice are proposed and proved. However, the question when the node set of PFS and that of concept lattice are the same is not solved yet, which we will go on to study.

Acknowledgement. This work is partially supported by the National Natural Science Foundation of China (NSFC) under Grant Nos. 61273019, 81273740, 81373767, 61201111, 61501397. It is also supported by Science Foundation of Hebei Province under Grant No. F2016203443. The authors gratefully acknowledge the supports. The authors are also grateful to the reviewers of the paper for their helpful comments and suggestions.

REFERENCES

[1] R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, *Proc. of the NATO Advanced Study Institute*, Banff, Canada, pp.445-470, 1982.
[2] B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundations*, Springer-Verlag, Berlin, 1999.
[3] T. Huysegoms, M. Snoeck, G. Dedene et al., Visualizing variability management in requirements engineering through formal concept analysis, *Procedia Technology*, vol.9, pp.189-199, 2013.
[4] V. Codocedo, I. Lykourentzou and A. Napoli, A semantic approach to concept lattice-based information retrieval, *Annals of Mathematics and Artificial Intelligence*, vol.72, no.1, pp.169-195, 2014.
[5] J. Baixerie, M. Kaytoue and A. Napoli, Characterizing functional dependencies in formal concept analysis with pattern structures, *Annals of Mathematics and Artificial Intelligence*, vol.72, no.1, pp.129-149, 2014.
[6] Z. Zhang, J. Du and L. Wang, Formal concept analysis approach for data extraction from a limited deep web database, *Journal of Intelligent Information Systems*, vol.41, no.2, pp.211-234, 2013.
[7] O. C. Cur, H. Maurer, N. H. Shah and P. L. Pendu, A formal concept analysis and semantic query expansion cooperation to refine health outcomes of interest, *BMC Medical Informatics & Decision Making*, vol.15, no.1, pp.1-6, 2015.
[8] J.-D. Kim and S.-H. Hwang, A FCA-based framework for discovering hidden knowledge from Twitter content, *International Journal of Software Engineering and Its Applications*, vol.10, no.1, pp.253-262, 2016.
[9] W. Hong, S. Li, J. Yu and J. Song, A new approach of generation of structural partial-ordered attribute diagram, *ICIC Express Letters, Part B: Applications*, vol.3, no.4, pp.823-830, 2012.
[10] J. Song, J. Yu, E. Yan, Z. Zhang and W. Hong, Syndrome differentiation of six meridians for warm disease based on structural partial-ordered attribute diagram, *ICIC Express Letters*, vol.7, no.3(B), pp.947-952, 2013.
[11] J. Yu, N. Chen, R. Sun, W. Hong and S. Li, Word sense disambiguation and knowledge discovery of English modal verb can, *ICIC Express Letters*, vol.7, no.2, pp.577-582, 2013.
[12] H. Meng, W. Hong, C. Zheng et al., Knowledge discovery of prescription for spleen deficiency syndrome base on attribute partial ordered structure diagram, *Proc. of the 2nd Information Technology and Mechatronics Engineering Conference*, Chongqing, China, pp.20-23, 2016.
[13] B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundations*, Science Press, Beijing, 2007 (in Chinese).