

The Connectivity of the Covering Approximation Space

Duixia Ma^(✉) and William Zhu

Lab of Granular Computing, Minnan Normal University, Zhangzhou, China
maduixia@163.com, williamfengzhu@gmail.com

Abstract. As a covering approximation space, its connectivity directly reflects a relationship, which plays an important role in data mining, among elements on the universe. In this paper, we study the connectivity of a covering approximation space and give its connected component. Especially, we give three methods to judge whether a covering approximation space is connected or not. Firstly, the conception of the maximization of a family of sets is given. Particularly, we find that a covering and its maximization have the same connectivity. Second, we investigate the connectivity of special covering approximation spaces. Finally, we give three methods of judging the connectivity of a covering approximation space from the viewpoint of matrix, graph and a new covering.

Keywords: Covering approximation space · Connectivity · Granular computing

1 Introduction

At the Internet age, data collected and stored are enormous and inexact. Then, there are many puzzles in data intelligence processing, such as how to take effective ways to cope with the uncertainty of ubiquitous information. So how to solve such bottleneck problem becomes an important issue in computer science and industry. In order to deal with this issue, researchers have developed many techniques such as rough set theory [9], fuzzy set theory [21], computing with words [13], and granular computing [8].

Rough set theory was proposed by Pawlak [10] in 1982 as a tool for dealing with the vagueness and granularity in information system. It is built on equivalence relation [11]. However, the data in powerful computer systems and storage media are complex and redundant in the information-based society. Thus equivalence relation imposes restrictions and limitations on many practical applications. Therefore, Pawlak's rough sets has been extended to the covering-based rough set theory [22, 23], relation-based rough sets [12, 19, 20], fuzzy rough sets [3, 18]. That is to say, Pawlak's approximation space is extended to generalization approximation space.

However, covering approximation space is a class of generalization approximation space. In addition, covering is a common data structure and is used

to describe overlapping information blocks in information systems. Therefore, it is important to investigate the covering approximation space. For a covering approximation space (U, \mathcal{C}) , where U is a universe and \mathcal{C} is a covering of U , it is called connected if x is connected to y for each pair $x, y \in U$, namely, there exist $K_1, K_2, \dots, K_n \in \mathcal{C}$ such that $x \in K_1, y \in K_n$ and $K_i \cap K_{i+1} \neq \emptyset$, for any $x, y \in U, i = 1, 2, \dots, n$. That is to say, the relationship between elements of universe relates to connectivity of this covering approximation space.

In this paper, we mainly investigate the connectivity of the covering approximation space and give three methods to judge its connectivity in terms of matrix, graph and a new covering. In order to study expediently, we firstly give a new covering, namely, the maximization of a covering. Particularly, a covering and its maximization are proved to have the same connectivity. So we study the connectivity of a covering approximation space through maximization of this covering. Second, we explore the connectivity of some special covering approximation spaces. Finally, we give three approaches to judge the connectivity of a covering approximation space by matrix, graph and a new covering. It is the core of this paper.

The remainder of this paper is organized as follows. In Sect. 2, we review some basic definitions and related conclusions about covering approximation space and graph theory. The conception of maximization of a covering and its properties are given in Sect. 3. Section 4 studies the connectivity of the special covering approximation spaces. The most important work is Sect. 5. This section gives three methods to judge the connectivity of a covering approximation space. Finally, we conclude this paper in Sect. 6.

2 Preliminaries

Covering is often used to describe overlapping information blocks in information systems. Graph theory is an intuitive and visible mathematical model. In this section, we introduce the basic definitions and related results about covering and graph theory.

2.1 Covering Approximation Space

In this subsection, we give the basic concepts of the covering approximation space.

Definition 1 (Covering [2]). *Let U be a universe of discourse, \mathcal{C} a family of subsets of U . If \mathcal{C} is a family of nonempty subsets of U , and $\cup \mathcal{C} = U$, \mathcal{C} is called a covering of U .*

Definition 2 (Covering approximation space [2]). *Let U be a nonempty set, \mathcal{C} a covering of U . The pair (U, \mathcal{C}) is called a covering approximation space.*

The minimal description is an important concept in a covering approximation space.

Definition 3 (Minimal description [24]). Let (U, \mathcal{C}) be a covering approximation space and $x \in U$. The family of sets $Md_{\mathcal{C}}(x) = \{C \in \mathcal{C} \mid x \in C \wedge (\forall S \in \mathcal{C} \wedge x \in S \wedge S \subseteq C \Rightarrow C = S)\}$ is called the minimal description of x . When there is no confusion, we omit the subscript \mathcal{C} .

In the same way, we give the definition of maximal description.

Definition 4 (Maximal description [14]). Let (U, \mathcal{C}) be a covering approximation space and $x \in U$. The family of sets $Maxd_{\mathcal{C}}(x) = \{C \in \mathcal{C} \mid x \in C \wedge (\forall S \in \mathcal{C} \wedge C \subseteq S \Rightarrow C = S)\}$ is called the maximal description of x . When there is no confusion, we omit the subscript \mathcal{C} .

In the following definitions, we introduce two types of coverings.

Definition 5 (Unary [25]). \mathcal{C} is unary if $|Md(x)| = 1$ for all $x \in U$.

Definition 6 (Pointwise-covered [25]). \mathcal{C} is called a pointwise-covered covering, if for any $K \in \mathcal{C}$ and $x \in K$, $K \subseteq \cup Md(x)$.

The connectivity of covering approximation space has been used in medical diagnosis [5]. We will introduce the connected covering approximation space.

Definition 7 [5]. Let (U, \mathcal{C}) be a covering approximation space.

- (1) Let $x, y \in U$. x is called to be connected to y if there are $K_1, K_2, \dots, K_n \in \mathcal{C}$ such that $x \in K_1, y \in K_n$ and $K_i \cap K_{i+1} \neq \emptyset$ for each $i = 1, 2, \dots, n - 1$.
- (2) (U, \mathcal{C}) is called connected if x is connected to y for each pair $x, y \in U$.

2.2 Graph Theory

Graph is an important tool at the area of mathematics. It has been employed to find rough set reducts [6]. In this subsection, we present the basic concepts of graph theory.

Definition 8 (Graph [17]). A set of elements and a relation between them is called a graph. Specifically, graph is a pair (V, E) , there V is called the vertex set of the graph, E is a unordered pair from the element of V , is called the edge set of the graph. We say that $G' = (V', E')$ is a subgraph of $G = (V, E)$ if $V' \subset V$ and $E' \subset E$.

Definition 9 [1]. (1) Chain(Walk): Let u and v are two vertexes of the graph. The chain of the graph is a series of finite vertexes and edges $u_0e_1u_1 \dots u_{n-1}e_nu_n$ ($u = u_0, v = u_n$), is called $u - v$ chain, where u_{i-1} and u_i , are placed adjacent to $e_i(1 \leq i \leq n)$, are two endpoint of the e_i .

(2) Path: The chain is called a path, if the internal point of chain is different.

(3) Connected graph: If there is a path between u and v , there u and v are two different vertexes of the graph, then the graph is connected, otherwise, the graph is not connected. The maximum connected subgraph is called a connected component.

Definition 10 (Bigraph [17]). A bigraph (or bipartite graph) $G = (V, E)$ is a graph whose vertices can be divided into two disjoint sets V_1 and V_2 such that every edge connects a vertex in V_1 to one in V_2 . We often use $G = (V_1 \cup V_2, E)$ to denote a bigraph.

3 Maximization of a Covering

Let \mathcal{C} be a family of sets on universe U . If there exist two elements K_1, K_2 of \mathcal{C} , such that K_1 is a subset of K_2 , then we omit K_1 , and so on. Like that, we can get some new sets \mathcal{C}' . Then the elements of \mathcal{C}' do not have this containment relationships. We call \mathcal{C}' the maximization of \mathcal{C} . First, we give a symbol of set theory [7].

Let U be a finite universe and \mathcal{A} be a family of subsets of U . Then $Max(\mathcal{A}) = \{X \in \mathcal{A} \mid \text{For all } Y \in \mathcal{A}. \text{ If } X \subseteq Y, \text{ then } X = Y\}$.

Definition 11 (*Maximization of sets*). Let \mathcal{C} be a set of sets. \mathcal{C}' is called the maximization of \mathcal{C} if $\mathcal{C}' = Max(\mathcal{C})$.

Example 1. Let $\mathcal{C} = \{\{a, c, d\}, \{a, b\}, \{c, d\}, \{f, d, b\}, \{a, b, f\}\}$ be a family of sets. Then the maximization of \mathcal{C} is that $\mathcal{C}' = \{\{a, c, d\} \{f, d, b\}, \{a, b, f\}\}$.

If a family of sets are a covering on the universe U , then the maximization of these sets is still a covering.

Proposition 1. Let $\mathcal{C} = \{K_1, K_2, \dots, K_n\}$ be a covering of the universe U and \mathcal{C}' the maximization of \mathcal{C} . Then \mathcal{C}' is also a covering of U .

Proof. Since \mathcal{C} is a covering of U , then $K_i \neq \emptyset (i = 1, 2, \dots, n)$ and $\bigcup_{i=1}^n K_i = U$. According to Definition 11, $K \neq \emptyset, \forall K \in \mathcal{C}'$. Since $\forall K_i \in \mathcal{C}$ there must exist $K_j \in \mathcal{C}'$, such that $K_i \subseteq K_j$, so $U = \bigcup_{i=1}^n K_i \subseteq \bigcup_{j \in I} K_j$, and $\bigcup_{j \in I} K_j \subseteq U$, then $\bigcup_{j \in I} K_j = U$. Therefore \mathcal{C}' is a covering of U .

The following proposition indicates that the process of maximizing can not change the connectivity of the covering approximation space.

Proposition 2. Let (U, \mathcal{C}) be a covering approximation space and \mathcal{C}' the maximization of \mathcal{C} . (U, \mathcal{C}) is connected if and only if (U, \mathcal{C}') is connected.

Proof. \Rightarrow : If (U, \mathcal{C}) is connected, then for any $x, y \in U, \exists K_1, \dots, K_m \in \mathcal{C}$ such that $x \in K_1, y \in K_m$ and $K_i \cap K_{i+1} \neq \emptyset (i = 1, 2, \dots, m)$. Since \mathcal{C}' the maximization of \mathcal{C} , thus there must exist $K'_1, \dots, K'_m \in \mathcal{C}'$ such that $K_i \subseteq K'_i$. Therefore $x \in K_1 \subseteq K'_1, y \in K_m \subseteq K'_m$ for any $x, y \in U$, and $K'_i \cap K'_{i+1} \supseteq K_i \cap K_{i+1} \neq \emptyset$. That is to say (U, \mathcal{C}') is connected.

\Leftarrow : It is straightforward.

Proposition 3. Let \mathcal{C} be a covering of the universe U and \mathcal{C}' the maximization of \mathcal{C} . Then $Max_{\mathcal{C}}(x) = Max_{\mathcal{C}'}(x)$ for all $x \in U$.

Proof. On one hand, for any $K \in Max_{\mathcal{C}}(x)$, then $x \in K$ and $\forall S \in \mathcal{C}, x \in S \wedge K \subseteq S \Rightarrow S = K$, thus $K \in \mathcal{C}'$. According to Definition 11, $S \not\subseteq K$ for all $S \in \mathcal{C}'$. Therefore, $K \in Max_{\mathcal{C}'}(x)$ i.e. $Max_{\mathcal{C}}(x) \subseteq Max_{\mathcal{C}'}(x)$. On the other hand, $\forall K' \in Max_{\mathcal{C}'}(x)$, then $\forall S \in \mathcal{C}, x \in S \wedge K' \subseteq S \Rightarrow S = K'$ based on Definition 11, that is to say $K' \in Max_{\mathcal{C}}(x)$, i.e. $Max_{\mathcal{C}}(x) \supseteq Max_{\mathcal{C}'}(x)$. To sum up, $Max_{\mathcal{C}}(x) = Max_{\mathcal{C}'}(x)$.

Proposition 4. *Let \mathcal{C} be a covering of the universe U and \mathcal{C}' the maximization of \mathcal{C} . Then $Md_{\mathcal{C}'}(x) = Maxd_{\mathcal{C}'}(x)$.*

Proof. $\forall K_1, K_2 \in \mathcal{C}'$, if $K_1 \neq K_2$, then $K_1 \not\subseteq K_2$, hence $Md_{\mathcal{C}'}(x) = Maxd_{\mathcal{C}'}(x)$.

4 The Connectivity of Covering Approximation Space

The connected covering approximation space has been used in medical diagnosis [5]. Therefore, it is important and necessary to study the connectivity of a covering approximation space.

Proposition 5. *Let (U, \mathcal{C}) be a covering approximation space and \mathcal{C}' the maximization of \mathcal{C} . If \mathcal{C}' is unary on U and $\mathcal{C}' \neq \{U\}$, then (U, \mathcal{C}) is not connected.*

Proof. We only need to prove that (U, \mathcal{C}') is not connected based on Proposition 2. If \mathcal{C}' is unary on U and $\mathcal{C}' \neq \{U\}$, then for all $x \in U$, $|Md(x)| = 1$. Thus we suppose (U, \mathcal{C}') is connected, then $\forall x, y \in U$, there exists $K_1, \dots, K_m \in \mathcal{C}'$ such that $x \in K_1, y \in K_m$ and $K_i \cap K_{i+1} \neq \emptyset$ ($1 \leq i \leq m - 1$). Let $x' \in K_1 \cap K_2$, thus $|Md(x')| = 2$. It is contradictory with $|Md(x)| = 1$ for all $x \in U$. Therefore (U, \mathcal{C}') is not connected. That is to say (U, \mathcal{C}) is not connected.

We only suppose that $\mathcal{C}' \neq \{U\}$ in the Proposition 5. It is because that \mathcal{C}' is connected when $\mathcal{C}' = \{U\}$.

Proposition 6. *Let (U, \mathcal{C}) be a covering approximation space. If \mathcal{C}' is unary on U , then $U|Md(x)$ ($x \in U$) is a connected component of (U, \mathcal{C}) .*

Proof. Since \mathcal{C}' is unary, then $Md(x) = Md(y)$ for any $y \in Md(x)$, thus $x \sim y$, and if $z \in U - Md(x)$, then $Md(x) \cap Md(z) = \emptyset$, so $x \not\sim z$. That is to say $U|Md(x)$ ($x \in U$) is a connected component of (U, \mathcal{C}) .

Proposition 7. *Let (U, \mathcal{C}) be a covering approximation space and \mathcal{C}' the maximization of \mathcal{C} . If \mathcal{C}' is a partition on U and $\mathcal{C}' \neq \{U\}$, then (U, \mathcal{C}) is not connected.*

Proof. Since \mathcal{C}' is a partition, thus $\forall K_1, K_2 \in \mathcal{C}'$, $K_1 \cap K_2 = \emptyset$, then $x \not\sim y$, for any $x \in K_1, y \in K_2$. Therefore (U, \mathcal{C}) is not connected.

Proposition 8. *Let (U, \mathcal{C}) be a covering approximation space and \mathcal{C}' the maximization of \mathcal{C} . If $\mathcal{C}' = \{K_1, \dots, K_m\}$ is a partition, then $U|K_i$ ($i = 1, 2, \dots, m$) is connected component of (U, \mathcal{C}) .*

Proposition 9. *Let (U, \mathcal{C}) , a covering approximation space, be connected and \mathcal{C}' the maximization of \mathcal{C} . Then there exists $x \in K$ for all $K \in \mathcal{C}'$, such that $|Md_{\mathcal{C}'}(x)| > 1$.*

Proof. If (U, \mathcal{C}') is connected, then we suppose that $|Md_{\mathcal{C}'}(x)| = 1$ for all $x \in K$. Since $Md_{\mathcal{C}'} = Maxd_{\mathcal{C}'}$, hence $\forall x \in K, \forall y \in U - K$, thus x is not connected y , which is contradictory with (U, \mathcal{C}') is connected. Therefore, there must exist $x \in K$ for all $K \in \mathcal{C}'$, such that $|Md_{\mathcal{C}'}(x)| > 1$.

The reverse of Proposition 9 may not hold from the following a counterexample.

Example 2. Let $U = \{a, b, c, d, e, f\}$ be a universe and $K_1 = \{a, b\}$, $K_2 = \{b, c\}$, $K_3 = \{f, d\}$, $K_4 = \{d, e\}$, $\mathcal{C} = \{K_1, K_2, K_3, K_4\}$. There exist $b \in K_1, K_2$ such that $|Md(b)| > 1$, and $d \in K_3, K_4$ such that $|Md(d)| > 1$, but (U, \mathcal{C}) is not connected.

5 Methods of Judging the Connectivity of a Covering Approximation Space

As a general covering approximation space, what we care is whether it is connected or not. In this section, we will explore the connectivity of the covering approximation space from the viewpoint of matrix, graph and a new covering.

5.1 From the Matrix Perspective

First, we give the matrix relating to the covering approximation space.

Definition 12. [15] Let (U, \mathcal{C}) be a covering approximation space and $U = \{u_1, u_2, \dots, u_n\}$, $\mathcal{C} = \{K_1, K_2, \dots, K_m\}$. We define a matrix $A = (a_{ij})_{m \times n}$ as follows:

$$a_{ij} = \begin{cases} 1, & u_j \in K_i, \\ 0, & u_j \notin K_i, \end{cases}$$

where $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$. Using u_j labels the j th column and K_i labels the i th row. A is called a matrix induced by \mathcal{C} .

Example 3. Let $U = \{a, b, c, d\}$ be a universe. $K_1 = \{a, b\}$, $K_2 = \{a, c, d\}$, $K_3 = \{b, c\}$, $K_4 = \{b, d\}$, $K_5 = \{c, d\}$, $\mathcal{C} = \{K_1, K_2, K_3, K_4, K_5\}$. Then $\mathcal{C}' = \{K_1, K_2, K_3, K_4\}$, hence the matrix A_{ij} as follows:

$$A_{ij} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

If we denote all column vectors of the A as $\mathcal{A} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$, then a family of vectors \mathcal{A} can be regarded as being composed of column vectors $\alpha_j (j = 1, 2, \dots, n)$.

Remark 1. Let $\alpha_j = (a_{1j}, a_{2j}, \dots, a_{mj})^\top$ be a column vector. Then we denote by $\|\alpha_j\| = a_{1j} + a_{2j} + \dots + a_{mj}$, where $j = 1, 2, \dots, n$.

Definition 13. Let A be a matrix induced by \mathcal{C} and \mathcal{A} a set of all column vectors of A . Then we denote the maximization of the \mathcal{A} by \mathcal{A}_{max} as follows:

$$\mathcal{A}_{max} = \{\alpha_i \in \mathcal{A} : \forall \alpha_j \in \mathcal{A}, \|\alpha_i\| \geq \|\alpha_j\|\}.$$

Example 4 (Continued from Example 3). $\mathcal{A}_{max} = \{ \alpha_2 \}$, where $\alpha_2 = (1\ 0\ 1\ 1)^\top$.

We take any one from \mathcal{A}_{max} , denoted as α . If a row of α is 1, then we omit column vectors whose values are 1 in the same row with α and the vectors we have omitted until there do not exist column vectors such that their values are 1 in the same row.

Example 5 (Continued from Example 3).

$$A_{ij} = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} \alpha_2 & \alpha_3 & \alpha_4 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} \alpha_2 & \alpha_4 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} \alpha_2 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

Since $\|\alpha_2\|$ is the maximization of $\|\alpha_i\| (i = 1, \dots, 4)$, hence we regard α_2 as α . Thus the first row of α is 1, so we omit α_1 whose the first row is 1. Because the values of α_3, α_4 and α_1 are 1 in the second row, we also omit α_3, α_4 .

Definition 14. Let (U, \mathcal{C}) be a covering approximation space and A be a matrix induced by \mathcal{C} . The new matrix through the above process is called the simplification of A , denoted by $Simp(A)$.

Proposition 10. Let (U, \mathcal{C}) be a covering approximation space, A be a matrix induced by \mathcal{C} and $N_{(U, \mathcal{C})}$ the number of connected component of (U, \mathcal{C}) . We denote the number of column vector of $Simp(A)$ by $N_{Simp(A)}$, then $N_{(U, \mathcal{C})} = N_{Simp(A)}$.

Proof. We suppose $A = (\alpha_1\ \alpha_2\ \dots\ \alpha_n)$, $Simp(A) = (\alpha_1\ \alpha_2\ \dots\ \alpha_m)$, $(m \leq n)$. On one hand, for any $\alpha_i (1 \leq i \leq n, i \neq 1, \dots, m)$ there exist $\alpha_j (j = 1, \dots, m)$ such that α_i, α_j have the same 1 in the same row, that is to say $a_i \sim a_j$ based on Definition 12. On the other hand, for any $\alpha_l, \alpha_k (l, k = 1, \dots, m)$, then $a_l \not\sim a_k$ based on Definition 14 and Definition 12. Therefore $N_{(U, \mathcal{C})} = N_{Simp(A)}$.

Proposition 11. Let (U, \mathcal{C}) be a covering approximation space, A be a matrix induced by \mathcal{C} . Then $Simp(A)$ only has one column vector if and only if (U, \mathcal{C}) is connected.

5.2 From the Graph Perspective

The following proposition is straightforward from the bigraph perspective [16].

Proposition 12. Let (U, \mathcal{C}) be a covering approximation space and G a bigraph induced by \mathcal{C} . (U, \mathcal{C}) is connected if and only if G is connected.

The above proposition gives a useful way to judge the connectivity of a covering approximation space. We will introduce another method through a general graph.

Definition 15 (Graph induced by K_i). Let $\mathcal{C} = \{K_1, K_2, \dots, K_m\}$ be a covering. We define the graph $G(K_i) = (V, E)$ induced by $K_i = \{x_1, x_2, \dots, x_l\}$ ($i = 1, \dots, m$) as follows:

- (1) $V = \{v_{K_i}, v'_{K_i}\};$
- (2) $E = \{x_1, x_2, \dots, x_l\}$, where x_j connects v_{K_i} with v'_{K_i} , $j = 1, 2, \dots, l$.

Example 6. Let $K = \{u_1, u_2, u_3, u_4\}$, then the graph induced by K as shown in Fig. 1:

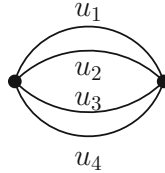


Fig. 1. The graph induced by K .

Remark 2. Let $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ be two graphs induced by K_1, K_2 respectively, where $V_1 = \{v_{K_1}, v'_{K_1}\}$, $V_2 = \{v_{K_2}, v'_{K_2}\}$. If $K_1 \cap K_2 \neq \emptyset$, then we denote as $v_{K_1} = v_{K_2}$ and $v'_{K_1} = v'_{K_2}$.

Definition 16 (Graph union [4]). Let $G(K_1) = (V_1, E_1)$, $G(K_2) = (V_2, E_2)$ be two graphs induced by K_1 and K_2 . Then the union of $G(K_1)$ and $G(K_2)$ is defined as a graph $(V_1 + V_2, E_1 + E_2)$, denoted by $G(K_1) \cup G(K_2)$.

It is clearly that a covering approximation space can induce a graph from the following definition.

Definition 17. Let (U, \mathcal{C}) be a covering approximation space. The graph induced by \mathcal{C} is called the graph induced by (U, \mathcal{C}) , denoted by $G(\mathcal{C})$.

Example 7. Let (U, \mathcal{C}) be a covering approximation space and $U = \{u_1, u_2, u_3, u_4, u_5\}$, $\mathcal{C} = \{\{u_1, u_3\}, \{u_3, u_5\}, \{u_2, u_4\}\}$. Then the graph induced by \mathcal{C} as shown in the Fig. 2:

Proposition 13. Let (U, \mathcal{C}) be a covering approximation space and $G(\mathcal{C})$ a graph induced by \mathcal{C} . $G(\mathcal{C})$ is connected if and only if (U, \mathcal{C}) is connected.

Proof. It is straightforward based on the Definitions 15 and 16 and Remark 2.

Proposition 14. Let (U, \mathcal{C}) be a covering approximation space and $G(\mathcal{C})$ a graph induced by \mathcal{C} . Then the numbers of connected component of (U, \mathcal{C}) and the graph $G(\mathcal{C})$ are equal.

Example 8 (Continued from Example 7). Since the graph induced by \mathcal{C} has two connected component from the Fig. 2, thus $N_{(U, \mathcal{C})} = 2$.

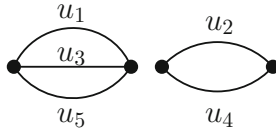


Fig. 2. The graph induced by \mathcal{C} .

5.3 From Covering Perspective

In the following discussion, we will continue to simplify a covering in terms of a covering for herself. First, we give the notion of friends of an element.

Definition 18 [26]. Let (U, \mathcal{C}) be a covering approximation space. For any $x \in U$, $\cup\{K : x \in K \wedge K \in \mathcal{C}\}$ is called the friends of x , denoted by $Friends(x)$.

According to Definition 18, it is clear that $\{Friends(x) : x \in U\}$ is a covering on the U , we denote by \mathcal{C}_1 . We already know that \mathcal{C}'_1 is also a covering and both of them have the same connectivity in Sect. 3. Now, for every element x on U , we can get the friends of x in the covering approximation space (U, \mathcal{C}'_1) . Then $\{Friends(x) : x \in U\}$ is also a covering, denoted by $\mathcal{C}_2 = \{Friends(x) : x \in U\}$, then we get \mathcal{C}'_2 . Following it until we get a partition \mathcal{C}'_N .

Example 9. Let (U, \mathcal{C}) be a covering approximation space. $\mathcal{C} = \{C_1, C_2, C_3\}$, $C_1 = \{a, c\}$, $C_2 = \{a, b, d\}$, $C_3 = \{b, e, f\}$. Since $Friends(a) = \{a, b, c, d\}$, $Friends(b) = \{a, b, d, e, f\}$, $Friends(c) = \{a, c\}$, $Friends(d) = \{a, b, d\}$, $Friends(e) = Friends(f) = \{b, e, f\}$. Thus $\mathcal{C}'_1 = \{\{a, b, c, d\}, \{a, b, d, e, f\}\}$. In the same way, we get $\mathcal{C}_2 = \{U, \{a, b, c, d\}, \{a, b, d, e, f\}\}$, then $\mathcal{C}'_2 = \{U\}$.

If \mathcal{C}'_N is a partition induced by above process, then this partition is called the simplification of \mathcal{C} . It is provided in the following Definition 19.

Definition 19. Let (U, \mathcal{C}) be a covering approximation space. If \mathcal{C}'_N is a partition induced by above process, then we call \mathcal{C}'_N is the simplification of \mathcal{C} , denote by \mathcal{C}'_N .

The notion of simplification of a covering can help us to judge the connectivity of a covering approximation space from the following proposition.

Proposition 15. Let (U, \mathcal{C}) be a covering approximation space and \mathcal{C}'_N the simplification of \mathcal{C} . Then (U, \mathcal{C}) is connected if and only if $\mathcal{C}'_N = \{U\}$.

Proof. It is straightforward based on the process of the simplification of a covering.

Proposition 16. Let (U, \mathcal{C}) be a covering approximation space and \mathcal{C}'_N the simplification of \mathcal{C} . Then an equivalence classes of \mathcal{C}'_N is a connected component of (U, \mathcal{C}) .

Proof. We suppose $\mathcal{C}'_N = \{K_1, \dots, K_m\}$, then $K_i \cap K_j = \emptyset (i, j = 1, \dots, m)$ based on Definition 19. According to the process of the simplification of a covering, there must exist a covering \mathcal{C}'_n and $z \in U$ such that $x \in Md_{\mathcal{C}'_n}(z)$ and $y \in Md_{\mathcal{C}'_n}(z)$ for any $x, y \in K_i (i = 1, \dots, m)$. Since (U, \mathcal{C}) and (U, \mathcal{C}'_n) have the same connectivity, thus $x \sim y$. On the other hand, for any $x \in K_i$ and $y \in K_j (i \neq j)$, since $K_i \cap K_j = \emptyset$, thus $x \not\sim y$. Therefore an equivalence classes of \mathcal{C}'_N is a connected component of (U, \mathcal{C}) .

6 Conclusions

In order to improve the high-efficiency in data mining, we have studied the connectivity of the covering approximation space and given three methods to judge its connectivity. First, we have given a conception of maximization of a family of sets, so the maximization of a covering has been given. Then we have investigated the relationship between covering and its maximization. Especially, we give that a covering and its maximization have the same connectivity. Second, we have studied the connectivity of special covering approximation spaces. Finally, we have given three methods to judge the connectivity of a covering approximation space with the aid of matrix, graph and a new covering.

Acknowledgments. This work is in part supported by The National Nature Science Foundation of China under Grant Nos. 61170128, 61379049 and 61379089, the Key Project of Education Department of Fujian Province under Grant No. JA13192, the Project of Education Department of Fujian Province under Grant No. JA14194, the Zhangzhou Municipal Natural Science Foundation under Grant No. ZZ2013J03, and the Science and Technology Key Project of Fujian Province, China Grant No. 2012H0043.

References

1. Bollobás, B.: *Modern Graph Theory*. Springer, New York (1998)
2. Bonikowski, Z., Bryniarski, E., Wybraniec-Skardowska, U.: Extensions and intentions in the rough set theory. *Inf. Sci.* **107**, 149–167 (1998)
3. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. *Int. J. Gen. Syst.* **17**, 191–209 (1990)
4. Gao, S.: *Graph Theory and Network Flow Theory*. Higher Education Press, Beijing (2009)
5. Ge, X.: Connectivity of covering approximation spaces and its applications on epidemiological issue. *Appl. Soft. Comput.* **25**, 445–451 (2014)
6. Jensen, R., Shen, Q.: Finding rough set reducts with ant colony optimization. In: *Proceedings of the 2003 UK Workshop on Computational Intelligence*, pp. 15–22 (2003)
7. Lai, H.: *Matroid Theory*. Higher Education Press, Beijing (2001)
8. Lin, T.Y.: Granular computing on binary relations. In: *Alpigrini, J.J., Peters, J.F., Skowron, A., Zhong, N. (eds.) RSCTC 2002. LNCS (LNAI), vol. 2475*, pp. 296–299. Springer, Heidelberg (2002)

9. Pawlak, Z.: *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, Boston (1991)
10. Pawlak, Z.: Rough sets. *Int. J. Comput. Inf. Sci.* **11**, 341–356 (1982)
11. Pawlak, Z., Skowron, A.: Rudiments of rough sets. *Inf. Sci.* **177**, 3–27 (2007)
12. Skowron, A., Stepaniuk, J.: Tolerance approximation spaces. *Fundamenta Informaticae* **27**, 245–253 (1996)
13. Wang, F.: Outline of a computational theory for linguistic dynamic systems: toward computing with words. *Int. J. Intell. Control Syst.* **2**, 211–224 (1998)
14. Wang, Z., Shu, L., Ding, X.: Minimal description and maximal description in covering-based rough sets. *Fundamenta Informaticae* **128**, 503–526 (2013)
15. Wang, S., Zhu, W., Zhu, Q., Min, F.: Characteristic matrix of covering and its application to boolean matrix decomposition. *Inf. Sci.* **263**, 186–197 (2014)
16. Wang, S., Zhu, W., Zhu, Q., Min, F.: Four matroidal structures of covering and their relationships with rough sets. *Int. J. Approximate Reasoning* **54**, 1361–1372 (2013)
17. West, D., et al.: *Introduction to Graph Theory*. Pearson Education, Singapore (2002)
18. Wu, W., Leung, Y., Mi, J.: On characterizations of (I, T) -fuzzy rough approximation operators. *Fuzzy Sets Syst.* **154**, 76–102 (2005)
19. Yao, Y.Y.: On generalizing pawlak approximation operators. In: Polkowski, L., Skowron, A. (eds.) *RSCTC 1998*. LNCS (LNAI), vol. 1424, p. 298. Springer, Heidelberg (1998)
20. Yao, Y.: Relational interpretations of neighborhood operators and rough set approximation operators. *Inf. Sci.* **111**, 239–259 (1998)
21. Zadeh, L.A.: Fuzzy sets. *Inf. Control* **8**, 338–353 (1965)
22. Zakowski, W.: Approximations in the space (u, π) . *Demonstratio Mathematica* **16**, 761–769 (1983)
23. Zhu, W., Wang, F.: Reduction and axiomization of covering generalized rough sets. *Inf. Sci.* **152**, 217–230 (2003)
24. Zhu, W.: Relationship between generalized rough sets based on binary relation and covering. *Inf. Sci.* **179**, 210–225 (2009)
25. Zhu, W., Wang, F.: On three types of covering-based rough sets. *IEEE Trans. Knowl. Data Eng.* **19**, 1131–1144 (2007)
26. Zhu, W., Wang, F.: A new type of covering rough sets. In: 2006 3rd International IEEE Conference on Intelligent Systems, pp. 444–449. IEEE (2006)