

Research Article

Communication P Systems on Simplicial Complexes with Applications in Cluster Analysis

Xiyu Liu and Alice Xue

School of Management Science and Engineering, Shandong Normal University, Jinan 250014, China

Correspondence should be addressed to Xiyu Liu, sdxyliu@163.com

Received 14 February 2012; Revised 16 April 2012; Accepted 17 April 2012

Academic Editor: M. De la Sen

Copyright © 2012 X. Liu and A. Xue. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The purpose of this paper is to propose a new kind of P systems on simplicial complexes. We present the basic discrete Morse structure, membrane structures on complexes, and communication rules. A new grid-based clustering technique is described based on this kind of new P systems. Examples are given to show the effect of the algorithm. The new P systems provide an alternative for traditional membrane computing.

1. Introduction

Membrane computing is a new branch of natural computing which is initiated by Păun et al. at the end of 1998, as an attempt to formulate models from the functioning of living cells [1], just like DNA computing coming from genes [2–4]. The advantage of these methods lies in its huge inherent parallelism which has drawn great attention from the scientific community so far. In recent years, many different models of P systems have been proposed, such as cell-like P systems, tissue-like P systems, and spiking neural P systems [5–9]. The obtained computing systems prove to be so powerful that they are equivalent with Turing machines [10] even when using restricted combinations of features and also computationally efficient. Up to now a number of applications were reported in several areas such as biology, biomedicine, linguistics, computer graphics, economics, approximate optimization, cryptography, and so forth.

Traditionally, Morse theory is the subject of differential topology and differential geometry where the topological spaces in question are smooth manifolds. When we want to study discrete problems, we will use combinatorial complexes rather than manifolds. Along this line, discrete Morse theory has been developed [11, 12]. Recently, discrete Morse theory has attracted many researchers because it has found applications in triangulations

and graphics. In fact, simplicial complex, the basic data structure in discrete Morse theory, will prove to be an important data structure besides trees and graphs.

Spatial cluster analysis is a traditional problem in knowledge discovery from databases [13]. It has wide applications since increasingly large amounts of data obtained from satellite images, X-ray crystallography, or other automatic equipment are stored in spatial databases. The most classical spatial clustering technique is due to Han and Kamber [13] who developed a variant PAM algorithm called CLARANS, while new techniques are proposed continuously in the literature aiming to reduce the time complexity or to fit for more complicated cluster shapes. Other clustering-like problems include impulsive cluster anticonsensus of discrete multiagent linear dynamic systems [14] and driving general complex networks into prescribed cluster synchronization patterns by using pinning control [15]. In another research, the authors introduce a cooperative article bee colony algorithm for solving clustering problems [16]. Also the authors propose a DNA-based clustering method by the Adleman-Lipton model [17]. For related research, one can also refer to [18–20].

In medical analysis there often appear data clustering problems of various type. The Wisconsin Breast Cancer Showhouse was founded in 1998 as an all-volunteer 501(c)(3) charitable organization by Nance Kinney, a breast cancer survivor (<http://www.breastcancershowhouse.org/wbcs2012/index.html>). Its mission is to support breast cancer and prostate cancer research at the Medical College of Wisconsin. This breast cancer databases were obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. Many computing methods have been applied to study this data case and, in fact, the Wisconsin Breast Cancer data set is becoming an important testing benchmark for soft computing.

Inspired by the above research, this paper focuses on the joint study of discrete Morse theory with membrane computing. Our purpose is to propose a P system on simplicial complex. Up to our knowledge, this is the first paper to extend membrane computing to complexes. Then we use membrane computing in cluster analysis, providing a new approach to data mining. We first propose a discrete Morse structure for a candidate of a class of new P systems. Then we described for the first time a communication P system on simplices. Then we propose a new method for cluster analysis by simplicial P systems. Finally, we present the Wisconsin Breast Cancer analysis.

2. Discrete Morse Structure

In this section we present some general discrete models which will form the basis of membrane structures. The main idea comes from [11, 21]. In order to do this, we need to present some basic topological concepts. For simplicity we always assume that we are working in an Euclidean space R^n .

2.1. Simplex without Orientation

A k -simplex (cell) σ is the convex hull of $k + 1$ affinely independent points. More precisely, suppose a_0, a_1, \dots, a_k are affinely independent, that is, $a_1 - a_0, \dots, a_k - a_0$ are linearly independent. Then σ is defined as the set of points in the form $x = \sum_{i=0}^k \lambda_i a^i$, where $\sum_{i=0}^k \lambda_i = 0$ and $\lambda_0, \lambda_1, \dots, \lambda_k \geq 0$. We will call k the dimension of a simplex and write $\dim(\sigma) = k$, while a_0, a_1, \dots, a_k are called vertices of the simplex. A simplex is uniquely indicated by its vertices

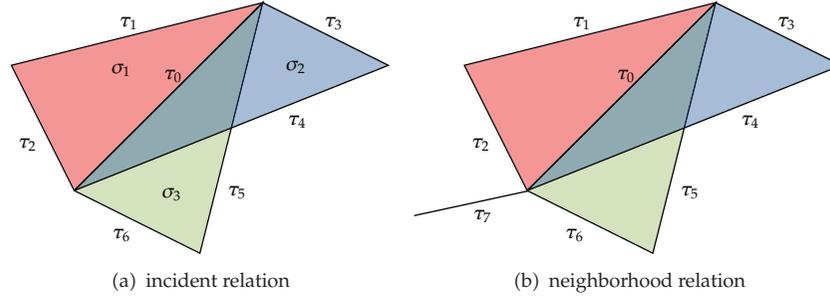


Figure 1: An attractor model on simplicial complex.

and hence is expressed as $\sigma = [a_0, a_1, \dots, a_k]$, or simply $\sigma = a_0 a_1 \cdots a_k$, and will be called a cell in this paper.

A face τ of a simplex σ is defined as a simplex generated by a nonempty subset of its vertices. We write $\tau < \sigma$. A face τ is called a *hyperface* of σ if $\dim \tau = \dim \sigma - 1$ and is denoted by $\tau < \sigma$. In this case, σ is called the parent of τ . Two cells τ_1 and τ_2 are called *incident* if $\tau_1, \tau_2 < \sigma$, and σ is called the *coface* of τ_1 and τ_2 . Two cells σ_1, σ_2 are called *neighbors* if they share a common hyperface. The cone from a vertex x to a k -simplex σ is the convex hull of x and σ which yields a $(k+1)$ -simplex $x\sigma$ provided x is not an affine combination of the vertices of σ . A *simplicial complex* K is a finite collection of nonempty simplices for which $\sigma \in K$ and $\tau < \sigma$ implies $\tau \in K$ and $\sigma_1, \sigma_2 \in K$ implies that $\sigma_1 \cap \sigma_2$ is either empty or a face of both. The underlying space of K is the union of simplices: $|K| = \bigcup_{\sigma \in K} \sigma$. K_p is a subset of K containing simplices of dimension p .

For a k -simplex σ , define K to be the collection of σ and all its faces. Then it is clear that K is a simplicial complex. We will call this simplicial complex a simple complex or, simply, a complex.

Now we consider some properties of incident and neighborhood relations as described above. First suppose $\tau_1, \tau_2 < \sigma$ are incident and $\dim \sigma \geq 2$. Then τ_1, τ_2, σ are faces of a simplex. Define $\psi = \tau_1 \cap \tau_2$ and we will show that $\psi \neq \emptyset$. This is evident because if $\sigma = v_{i_0} v_{i_1} \cdots v_{i_p}$, then $p \geq 2$. By removing one vertex we obtain τ_1 and τ_2 with p vertices remained. Since $p \geq 2$ we know that there exists at least one common vertex among τ_1, τ_2 . By the definition of K we get $\psi \in K$ and consequently τ_1, τ_2 are neighbors.

Conversely, if $\tau_1, \tau_2 > \psi$ are neighbors, they need not be incident as shown in Figure 1. In the case when they are in the same simplex, however, this is true. In fact, suppose $\psi = v_{i_1} \cdots v_{i_p}$. Then all the vertices $v_{i_1} \cdots v_{i_p}$ belong to τ_1 and τ_2 . Then we can assume that $\tau_1 = v_{i_1} \cdots v_{i_p} v_{j_1}$ and $\tau_2 = v_{i_1} \cdots v_{i_p} v_{j_2}$. Since both cells are in the same simplex, define $\sigma = v_{i_1} \cdots v_{i_p} v_{j_1} v_{j_2}$. Then $\sigma \in K$. Consequently τ_1, τ_2 are incident. Putting everything together we get a theorem.

Theorem 2.1. *Two incident cells τ_1, τ_2 are neighbors if they are at least one-dimensional. Two neighboring cells are incident provided they are located in the same simplex.*

By the definition of neighborhood, if σ_1, σ_2 are neighbors and their common hyperface is τ , then clearly $\tau = \sigma_1 \cap \sigma_2$ and, consequently, this common hyperface is unique. However, there exist cells with nonempty intersection but they are not neighbors. Now we consider incident cells $\tau_1, \tau_2 < \sigma$. We will show that their coface is also unique. First if $\dim \tau_1 = \dim$

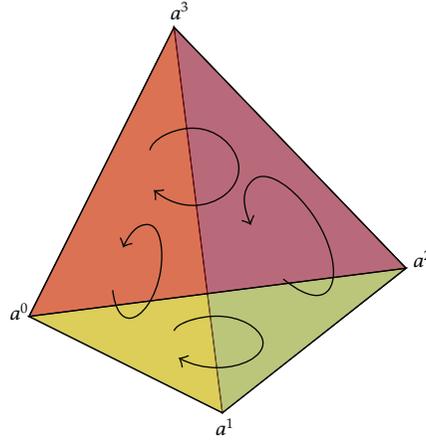


Figure 2: A simplicial complex with orientation in R^3 .

$\tau_2 = 0$, then $\dim \sigma = 1$. Therefore σ is the edge joining the two vertices τ_1, τ_2 and hence is unique. Next if $\dim \tau_1 = \dim \tau_2 \geq 1$, then as described previously we define $\psi = \tau_1 \cap \tau_2$. Then $\tau_1 = v_1\psi, \tau_2 = v_2\psi$ and $v_1 \neq v_2$. Therefore $\sigma = v_1v_2\psi$ and is unique.

Theorem 2.2. *Neighboring cells can only share a unique common hyperface. Incident cells have a unique coface (parent).*

2.2. Simplex with Orientation

Simplicial complexes with orientation are important tools in the study of topological properties of discrete data structure. Concepts about discrete Morse functions are listed as follows (Robin [11]). For a simplex $\sigma = [a_0, a_1, \dots, a_k]$, there are two orientations and the opposite orientation is denoted by $-\sigma$. Figure 2 shows the orientation of a three-dimensional complex.

In the following we will use $[a_0, a_1, \dots, \hat{a}_i, \dots, a_k]$ to denote a face of σ , where the vertex a_i is eliminated. The following chain is defined as the boundary of the complex:

$$\partial\sigma = \sum_{i=0}^k (-1)^i [a_0, a_1, \dots, \hat{a}_i, \dots, a_k]. \quad (2.1)$$

If g_i are integers, then $\sum_i g_i \sigma_i$ is called a chain, where $\dim \sigma_i$ remains the same within a chain. Boundary operator extends to chains naturally. An important property of boundary operator is that $\partial \circ \partial = 0$. For a $(k-1)$ -dimensional simplex τ^{k-1} and a k -dimensional simplex σ_k , define its relationship operator as follows:

$$[\sigma^k : \tau^{k-1}] = \begin{cases} 0, & \tau \not< \sigma \\ 1, & \tau < \sigma \text{ with same orientation} \\ -1, & \tau < \sigma \text{ with opposite orientation.} \end{cases} \quad (2.2)$$

In simplicial complex we can define Morse functions which is a tool for optimization.

Definition 2.3. Let K be a simplicial complex. A function $f : K \rightarrow R$ is a discrete Morse function if for every $\sigma^p \in K_p$ the following two statements are true:

$$(1) \#\{\tau^{p+1} > \sigma \mid f(\tau) \leq f(\sigma)\} \leq 1; \#\{v^{p-1} < \sigma \mid f(v) \geq f(\sigma)\} \leq 1.$$

A simplex σ^p is critical with index p if

$$(2) \#\{\tau^{p+1} > \sigma \mid f(\tau) \leq f(\sigma)\} = 0; \#\{v^{p-1} < \sigma \mid f(v) \geq f(\sigma)\} = 0.$$

Discrete gradient can also be defined on the complex K . If v is a critical point, then define $V(v) = 0$. Otherwise if $v \in K_0$ is not critical and the edge e satisfies $e > v$, $f(e) \leq f(v)$, then define $V(v) = \pm e$ where the sign is determined so that $\langle \partial(V(v)), v \rangle = -1$. Here the inner product $\langle \cdot, \cdot \rangle$ is the obvious inner product on oriented chains with respect to which the oriented simplices are orthonormal. It is easy to see that, if the edge $e = vu$, then $V(v) = -e$. Generally speaking, discrete gradient is a mapping

$$V : C_p(K, Z) \longrightarrow C_{p+1}(K, Z) \quad (2.3)$$

A natural gradient flow is $\Phi(\sigma) = \sigma + \partial V(\sigma) + V(\partial\sigma)$.

3. Communication P Systems on Simplices

3.1. Traditional P Systems

Membrane is a structure serving as a protected reactor. We will identify a membrane m with its delimited space. When we say inclusion for membranes, it is always strict inclusion. Now we list some elementary concepts concerning the basic operations of membranes:

- (i) m, m' are vicinal, if $m' \subset m$ and there is no m'' such that $m' \subset m'' \subset m$,
- (ii) *elementary membrane*: with no lower vicinal membranes, *skin membrane*: with no upper vicinal membranes, we assume there is always a unique skin,
- (iii) *degree*: number of membranes,
- (iv) *sibling membranes* m, m' : if there is a m'' which is upper vicinal for both m and m' .

Parentheses expression is often used to describe membrane structures. For example, the membrane structure as shown in Figure 3 has a parentheses expression for membranes as follows:

$$[[[[_2[[[_7[[[_5[[[_8[[[_9]]_6]]_4]]_3]]_1]. \quad (3.1)$$

For a set U , a *multiset* over U is a mapping $M : U \rightarrow N$, where N is the set of nonnegative integers. For $a \in U$, $M(a)$ is the multiplicity of a in M . Suppose the set of objects is O with a subset E such that objects from E are available in the environment in arbitrary multiplicities, that is, its multiset is $M : E \rightarrow \{\infty\}$. A P system with symport/antiport rules of degree $m \geq 1$ is a construct

$$\Pi = (O, T, E, \mu, \omega_1, \dots, \omega_m, R_1, \dots, R_m, i_0), \quad (3.2)$$

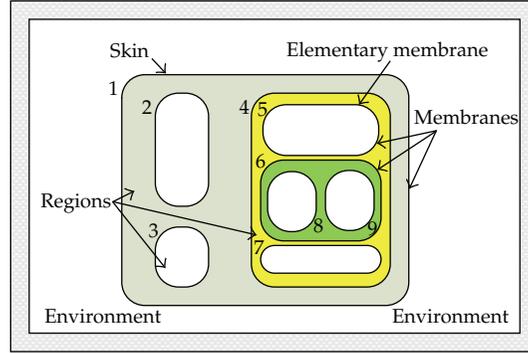


Figure 3: Traditional membrane structures.

where O is the alphabet, $T \subset O$ is the alphabet of terminal objects, μ is a membrane structure of degree m , $\omega_1, \dots, \omega_m$ are the multisets of objects associated with the m regions of μ , and R_1, \dots, R_m are finite sets of symport and antiport rules associated with the m membranes of μ , and i_0 is the input/output region.

A P system is called *stable* if, even if some rules are still applicable, their application does not change the string/object content of the membrane structure, nor the membrane structure itself. For a subalphabet $W \subset O$, we call a system stable over W if the projection over W of the string/object remains unchanged, even if some rules are still applicable. If $W = \{a\}$, we will say stable over a . If $R = (R'_1, \dots, R'_m)$, with $R'_i \subset R_i$ for $1 \leq i \leq m$, is a subset of rules, we call a P system stable with respect to the rules R' if the P system with rules R' is stable (i.e., applications of rules in R' do not change the string/object content of the system's membranes) [19].

3.2. Membrane Structures on Simplices

Now we describe membrane structures on simplicial complexes. First we assume that the complex is a simple complex, that is, a simplex with all its faces including vertices. In Figure 4 a simple three dimensional example is presented with 15 membranes. In the general case when there is a complex $\{a^0, a^1, \dots, a^n\}$ in R^n , the number of simplices are

$$1 + C_{n+1}^n + C_{n+1}^{n-1} + \dots + C_{n+1}^1 = \sum_{i=0}^{n+1} C_{n+1}^i - 1 = 2^{n+1} - 1. \quad (3.3)$$

The boundary relations of simplices are shown in Figure 5 where the arrows point to the boundary cells.

Now we consider the general cases where K is a simplicial complex in R^n . For an example of K in R^3 as shown in Figure 6, the total number of cells is $6 + 19 + 21 + 9 = 55$. For example, the three dimensional simplices are listed as

$$a^0 a^1 a^4 a^3, \quad a^3 a^1 a^4 a^2, \quad a^3 a^4 a^2 a^7, \quad a^1 a^5 a^2 a^4, \quad a^4 a^5 a^6 a^2, \quad a^4 a^6 a^7 a^2. \quad (3.4)$$

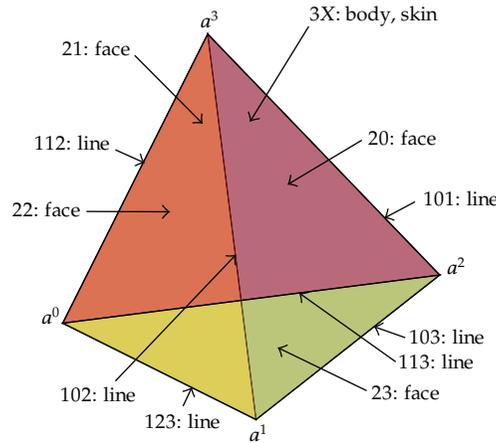


Figure 4: A simplex membrane structure.

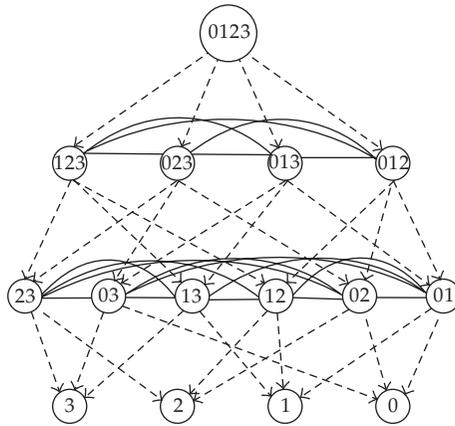


Figure 5: A network model for membrane structure.

Generally speaking, a simplicial complex K is denoted by a set of vertices $\{a^0, a^1, \dots, a^k\}$ in R^n . A simplex $\sigma \in K$ is called membrane. A membrane is called a max-simplex if it is not a face of another simplex in K . A simplex is denoted by its vertices. Evidently, vertices are zero-dimensional cells and hence are elementary membranes. If $\tau < \sigma$ is a face of σ , then we say that σ is the parent of τ . Figure 5 shows the parents and neighborhood relations of the complex in Figure 4. For the more general simplicial complex as in Figure 6, the network model is shown in Figure 7.

If σ_1, σ_2 are incident, we say there is an upper link (channel) between σ_1, σ_2 . If σ_1, σ_2 are neighbors, we say there is a lower link between them. An upper link is denoted by $(\sigma_1, \sigma_2)^\tau$, where τ is their (common) parent, while lower link is written as $(\sigma_1, \sigma_2)_\tau$, where τ is their common hyperface. Upper link is also written by $(\sigma_1, \sigma_2)_U$ or, simply, (σ_1, σ_2) , while lower link is denoted by $(\sigma_1, \sigma_2)_L$. Links have no directions. Thus (σ_1, σ_2) and (σ_2, σ_1) are identical. We will specify one from these two links, and only one is allowed.

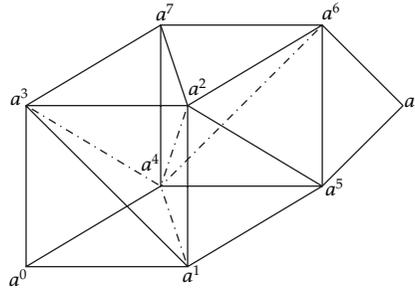


Figure 6: A simplicial complex model with all cells.

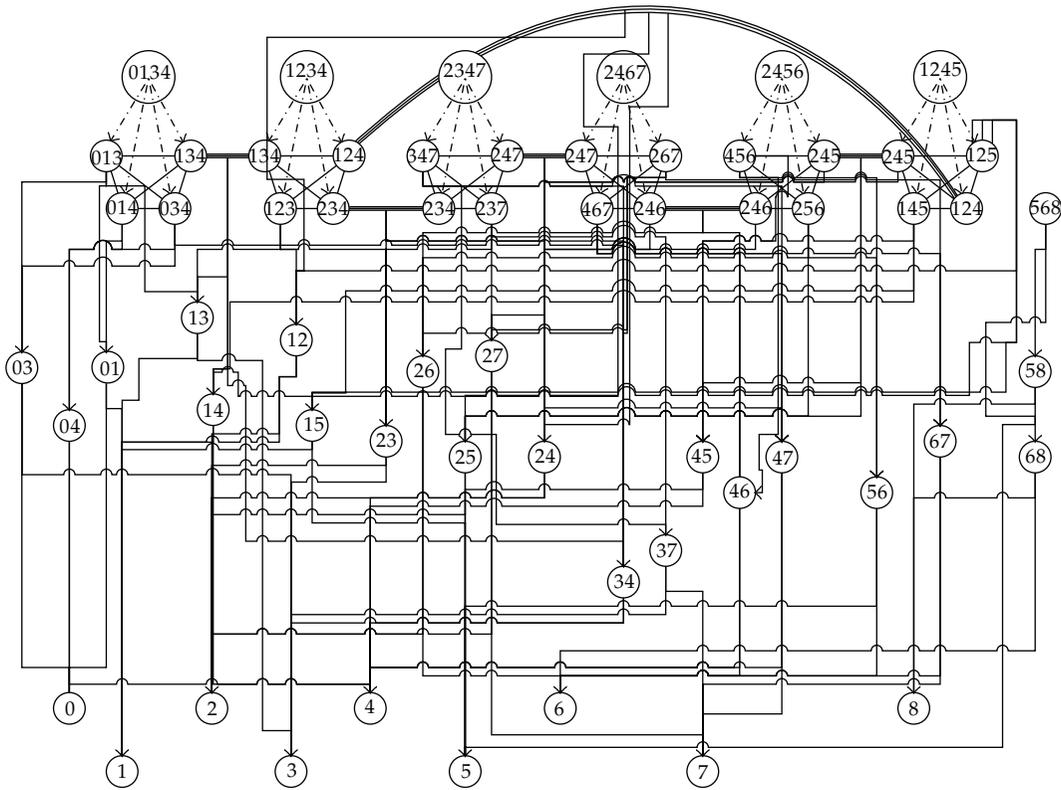


Figure 7: A network model for the simplicial complex model in Figure 5.

Definition 3.1. A P system on a simplicial complex K , called a simplicial P system, with antiport and symport rules is a construct

$$\Pi = \left(m, O, E, \omega_1, \dots, \omega_m, R_1, \dots, R_m, \text{ch}, F(i, j) \Big|_{(i,j) \in \text{ch}}, i_0 \right), \quad (3.5)$$

where m is the number of cells labeled with $1, 2, \dots, m$, O is the alphabet, E is the set of objects with unlimited multiplicity in the environment, $\omega_1, \dots, \omega_m$ are initial strings over O of multiset, R_1, \dots, R_m are symport and antiport rules associated with the m membranes,

$\text{ch} \subseteq \{(i, j)_R, (i, j)_L : i, j \in \{1, \dots, m\}\}$ is the set of links, and $F(i, j)$ is a finite set of antiport and/or symport rules associated with the link $(i, j) \in \text{ch}$.

Ceterchi and Martin-Vide [19] proposed a new type of communication P systems with priority relations. They introduced a promoter for a rule to be active and a inhibitor for it to be inactive. Induced by their idea, we will present an ordered system in this section.

An antiport rule $(u, \text{out}; v, \text{in})_i$ in R_i exchanges the multiset u inside R_i with v outside it. A symport rule $(x, \text{out})_i$ (or $(x, \text{in})_i$) sends out (takes in) the multiset x with respect to membrane R_i . For a specific membrane R_i , rules are totally ordered as $r_1 > r_2 > \dots > r_n > r_{n+1} > \dots$. The rule r_{n+1} is applicable if and only if the system has reached a stable configuration with respect to rules $r_1 > \dots > r_n$. We can use a queue structure to represent this process.

3.3. Rules in Simplicial P Systems

Now we describe the communication rules in simplicial P systems. For our purpose in this paper, there are mainly four types of communication rules in a simplicial P system. Each type of rule may have operators such as out, in, up, down.

First suppose $(\tau_1, \tau_2)_\sigma$ is an upper link. A rule like $[(x, y), \text{up}]^U$ means that the multiset x and y from τ_1 and τ_2 transform into z and go up to their parent σ . For two cells $\tau < \sigma$, the antiport rule $[u, \text{out}; v, \text{in} \mid \tau < \sigma]$ in τ means exchanging multiset u inside membrane τ with the multiset v outside it (in σ). The symport rule $[u, \text{out} \mid \tau < \sigma]$ sends the multiset u outside τ . Another symport rule $[u, \text{in} \mid \tau < \sigma]$ works similarly.

An upper link rule in $R^\sigma(\tau_1, \tau_2)$ may have the following forms:

$$[(x, y), \text{up}; (\alpha, \beta), \text{in}]^U \longrightarrow [z, \text{in}; \gamma, \text{down}], \quad (3.6)$$

$$[u, \text{out}; v, \text{in} \mid \tau_1 < \sigma], \quad [u, \text{out}; v, \text{in} \mid \tau_2 < \sigma]. \quad (3.7)$$

An lower link rule in $R_\sigma(\tau_1, \tau_2)$ may have the following forms:

$$[(x, y), \text{down}; (\alpha, \beta), \text{in}]_L \longrightarrow [z, \text{in}; \gamma, \text{up}], \quad (3.8)$$

$$[u, \text{out}; v, \text{in} \mid \sigma < \tau_1], \quad [u, \text{out}; v, \text{in} \mid \sigma < \tau_2]. \quad (3.9)$$

Equation (3.6) means moving x from cell τ_1 and y from τ_2 to z to parent σ and simultaneously moving γ from σ to α into τ_1 and β into τ_2 . Equation (3.8) means moving x from cell τ_1 and y from τ_2 to z to hyperface σ and simultaneously moving γ from σ to α up to τ_1 and β up to τ_2 .

3.4. Configuration and Computation

Now we describe the configuration and computation of simplicial P systems. For our purpose, change of membrane structure is not involved. A *configuration* of a simplicial P system is the state of the system described by specifying the objects and rules associated to each membrane. The initial state is called *initial configuration*. Therefore, the multisets represented by $\omega_i, 1 \leq i \leq d$ in Π , constitute the initial configuration of the system.

The system evolves by applying rules in membranes and this evolution is called *computation*. The computation starts with the multisets specified by $\omega_1, \dots, \omega_m$ in the m cells. In each time unit, rules are used in a cell. If no rule is applicable for a cell, then no object changes in it. The system is synchronously evolving for all cells.

When the system has reached a configuration in which no rule is any longer applicable, we say that the computation *halts*. A configuration is *stable* if, even if some rules are still applicable, their application does not change the object content of the membranes. The computation is successful if and only if it halts, or it is stable. The result of a halting/stable computation is the number described by the multiplicity of objects present in the cell i_0 in the halting/stable configuration.

4. Cluster Analysis by Simplicial P Systems

4.1. Problem Setting and Algorithm

Now suppose the data set to be clustered is $X = \{z_1, z_2, \dots, z_N\} \subset \prod_{i=1}^n [P_i, Q_i] \subset R^n$. We now construct a uniform grid in R^n as follows. Choose integers $n_i > 0$ and let $\delta_i = (Q_i - P_i)/n_i$, and divide the interval $[P_i, Q_i]$ into n_i subintervals $[P_i, P_i + \delta_i), \dots, [Q_i - \delta_i, Q_i)$. Therefore X is contained in the following grid \mathcal{G} with $N = n_1 \cdot n_2 \cdot \dots \cdot n_n - N_0$ cells, where N_0 is the number of cells that do not contain data:

$$\bigcup_{t_1=1}^{n_1} \dots \bigcup_{t_n=1}^{n_n} \prod_{i=1}^n [P_i + (t_i - 1)\delta_i, P_i + t_i\delta_i). \quad (4.1)$$

Then we can define a new data set $\Omega = \{x_1, \dots, x_N\}$, where x_i is each grid point $x_i = [x_i^1, \dots, x_i^n]$:

$$x_i^j = P_j + (t_j - 1)\delta_j, \quad 1 \leq j \leq n, \quad 1 \leq t_j \leq n_j. \quad (4.2)$$

Define a weight function on Ω by $m(x) = \#\{z_i \in \text{the cell corresponding to the data } x\}$. In this way, the original data set is transformed into a new data set Ω with weights equal to the original number of points. Next we will always work on the data set Ω . For simplicity, we will consider a density-based clustering technique. For two points $x, y \in \Omega$, the similarity is defined by

$$f(x, y) = \frac{\min\{m(x), m(y)\}}{d(x, y)}, \quad (4.3)$$

where $d(x, y)$ is the topological distance of the two points. That is, if x, y are incident, then $d(x, y) = 1$. Else $d(x, y)$ is the minimal number of edges which form a path connecting x and y . For two subsets $C_1, C_2 \subset X$, the similarity is defined by

$$f(C_1, C_2) = \max_{x \in C_1, y \in C_2} \{f(x, y) \mid x, y \text{ are incident}\}. \quad (4.4)$$

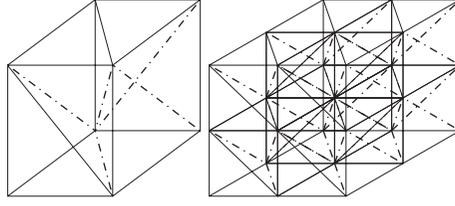


Figure 8: Simplices constructed for three dimensional grids.

The clustering is implemented by a hierarchical method as follows. At first, each point of the data set X forms a cluster which contains a singleton. Then each data point tries to connect another data point in its neighborhood. After this step, a cluster can be found as connected points. Now we construct a simplicial complex K corresponding to the data set Ω . On the basis of the rectangles as in (4.1), we add some hyperplanes to form a triangulation. Then each rectangle is decomposed as several simplices. Hyperplanes can be chosen such that the set of simplices satisfy the definition of simplicial complex. And then K is defined as the union of such simplices. In the three dimensional case, this is shown as in Figure 8.

Now we will show that the triangulation as above exists. In fact, we need only to consider an inner cell $C = \Pi_{i=1}^n [p_i, q_i]$, where $r_i < p_i < q_i < s_i$. There are 2^n vertices for this cell

$$\left\{ [x^1, \dots, x^n] \mid x^i = p_i, q_i \right\}. \quad (4.5)$$

There are totally $3^n - 1$ surrounding cells

$$\left\{ \prod_{i=1}^n [x_i, y_i] \mid [x_i, y_i] = [r_i, p_i), [p_i, q_i), [q_i, s_i) \right\} \setminus \{C\}. \quad (4.6)$$

First we consider the triangulation of the cell C . Now we choose one vertex $v_0 = [z_1, \dots, z_n]$ from the cell C where z_i can be p_i or q_i . Now we denote

$$\bar{y}_i = \begin{cases} p_i, & \text{if } y_i = q_i \\ q_i, & \text{if } y_i = p_i. \end{cases} \quad (4.7)$$

Consider the vertex set

$$S_j = \left\{ [x^1, \dots, x^{j-1}, \bar{y}^j, x^{j+1}, \dots, x^n] \mid x^i = p_i, q_i \right\}. \quad (4.8)$$

Then $\#(S_j) = 2^{n-1}$. Clearly C can be decomposed as disjoint cones $C = \bigcup_{j=1}^n v_0 S_j$. Notice that S_j is an $(n-1)$ -dimensional rectangle and this means that, for any triangulation of S_j , we can join them with v_0 to form a triangulation of $v_0 C$. Therefore by induction we get the construction of triangulation.

Lemma 4.1. *The decomposition $C = \bigcup_{j=1}^n v_0 S_j$ is valid with the property that each pair of $\text{Int}(v_0 S_j)$ is disjoint, where $\text{Int}(C)$ is the set of interior points in R^n .*

Inputs: $\Omega = \{x_1, x_2, \dots, x_N\}$, δ_0 : similarity threshold value, $\mathcal{N}_i = \{y \mid y, x_i \text{ are incident}\}$
Outputs: $\mathcal{C} = \{C_1, C_2, \dots, C_t\}$: set of clusters, t : number of clusters, Ω_0 : outliers
Begin
Set $\Omega_1 = \Omega_0 = \emptyset$, $C_1 = \emptyset$, $t = 1$.
for $i = 1$ **to** N
 If $\mathcal{N}_i = \emptyset$, then add x_i into set Ω_0 . Otherwise add x_i into set Ω_1 . For each $x_j \in \mathcal{N}_i$, $j > i$, calculate the similarity measure $f(x_i, y)$. If $f(x_i, y) > \delta_0$, then set the edge $\sigma(x_i, y)$ as active. In this case, set $C_1 = \{x_i\}$ if C_1 is still an empty set.
end
while $\Omega_1 \neq \emptyset$ **do**
 (1) For each $x \in \Omega_1$ if there exists an active edge $\sigma(xy)$ with $x \neq y$, $y \in C_t$, then add x into C_t and remove x from Ω_1 .
 (2) $t = t + 1$.
end
End

Algorithm 1: A self-joining clustering algorithm.

Proof. Suppose $\mathbf{x} = [a_1, \dots, a_n] \in C$ and $x \neq v_0$. Consider the point $\mathbf{x}_t = (1-t)v_0 + t\mathbf{x} = [(1-t)z_1 + ta_1, \dots, (1-t)z_n + ta_n]$ where $t \geq 1$. Let t_0 be the first zero point of the following function:

$$\psi(t) = \min_{1 \leq j \leq n} [\bar{y}_j - (1-t)z_j - ta_j]. \quad (4.9)$$

Clearly $\psi(1) \geq 0$. Let $[\bar{y}_j - (1-t_0)z_j - t_0a_j] = 0$. Then $\mathbf{x} \in v_0S_j$. □

By the above discussion, we also have another lemma.

Lemma 4.2. *Suppose the triangulation of S_j in R^{n-1} is $\{T_j^s\}$. Then the set $\cup_{j=1}^n \cup_s \{v_0T_j^s\}$ forms a triangulation of C .*

Finally we obtain a theorem.

Theorem 4.3. *A simplicial complex K exists corresponding to the grids \mathcal{G} .*

For each node $x \in \Omega$, define its neighborhood as $\mathcal{N}(x) = \{y \in \Omega \mid y \text{ is incident with } x \text{ in } K\}$. For a subset $C \subset \Omega$, define its neighborhood $\mathcal{N}(C)$ as $\mathcal{N}(C) = \cup_{x \in C} \mathcal{N}(x)$.

Now we propose an algorithm for our clustering problem. This is a self-joining cluster technique. Initially, one data point in Ω forms a cluster of singleton. Then each node searches for its neighborhood. If there exists a neighboring node which is similar enough, then activate a link between the two nodes. The final cluster is linked nodes. To define the meaning of similar enough, we need a parameter δ_0 such that $f(x, y) > \delta_0$. Putting everything together, we get the algorithm as shown in Algorithm 1.

4.2. Design of a P System

Now we have already defined the simplicial complex K as the membrane structure with M total membranes. Next we need to specify the alphabet and rules to be used. First we design a binary coding scheme for the weight function $m(\cdot)$ and the distance function $d(x, y)$. Suppose

the length of the coding is L . In this way, the weight function and the distance function take binary strings as values. Now we suppose $m(x_i) = m_{i1} \cdots m_{iL}$, where $m_{ij} = 0, 1$. We also need an integer H such that if the coding of δ_0 is $d_1 d_2 \cdots d_L$, then

$$\sum_{i=1}^L d_i 2^{L-i} = H. \quad (4.10)$$

Suppose $\mu \subset \text{ch}$ is a subset of links. Then we can define a P system as follows:

$$\Pi = (O, \mu, \omega_1, \dots, \omega_M, R(i, j)). \quad (4.11)$$

The working alphabet is

$$\begin{aligned} O = & \{a_{is}, b_{is} : 1 \leq i \leq N, 1 \leq s \leq L\} \\ & \cup \{g_{is}, h_{is} : 1 \leq i \leq N, 1 \leq s \leq L\} \\ & \cup \{\alpha, \beta, \beta_+, \beta_-\} \cup \{s_{ij} : 1 \leq i, j \leq N\}. \end{aligned} \quad (4.12)$$

Let $v_i = x_i$, $i = 1, \dots, N$ be the vertices. We will use $\sigma(v_{i_0} \cdots v_{i_k})$ to denote a k -dimensional simplex. Therefore $\sigma(v_i)$ is a vertex or v_i for short. Define $N_i = \#(\mathcal{N}_i)$. ch is the set of edges in K .

The initial multiset ω_i stands for the multiset in the membrane $\sigma(v_i)$:

$$\begin{aligned} \omega_i = & \{a_{is}^{2^{L-s} N_i}, b_{is}^{2^{L-s} N_i} : 1 \leq s \leq L\} \\ & \cup \{g_{is}^{m_{is}}, h_{is}^{1-m_{is}} : 1 \leq s \leq L\} \cup \{\beta, \beta_+\}. \end{aligned} \quad (4.13)$$

For $(i, j) \in \mu$, rules are $R(i, j) = \{A_{ij}^s, B_{ij} : s = 1, \dots, L\}$ with the order as follows:

$$\begin{aligned} & A_{ij}^1 < A_{ij}^2 < \cdots < A_{ij}^L, \quad i \cdot j = 1, \dots, N \\ A_{ij}^s = & \{\beta^2 g_{is} h_{js} \rightarrow (\sigma(v_i v_j), \beta_+)\} \\ & \cup \{\beta^2 h_{is} g_{js} \rightarrow (\sigma(v_i v_j), \beta_-)\} \\ & \cup \{(\sigma(v_i v_j), \beta_+) \rightarrow (\sigma(v_i), \beta_+^{2^L})\} \\ & \cup \{(\sigma(v_i v_j), \beta_-) \rightarrow (\sigma(v_j), \beta_-^{2^L})\} \\ B_{ij} = & \{\beta_+ a_{is} \rightarrow (\sigma(v_i v_j), s_{ij}), \text{in} : 1 \leq s \leq L\} \\ & \cup \{\beta_- a_{js} \rightarrow (\sigma(v_i v_j), s_{ij}), \text{in} : 1 \leq s \leq L\}. \end{aligned} \quad (4.14)$$

Rules on edges are

$$R_{ij} = \{s_{ij}^H \rightarrow \alpha : 1 \leq s \leq L\}. \quad (4.15)$$

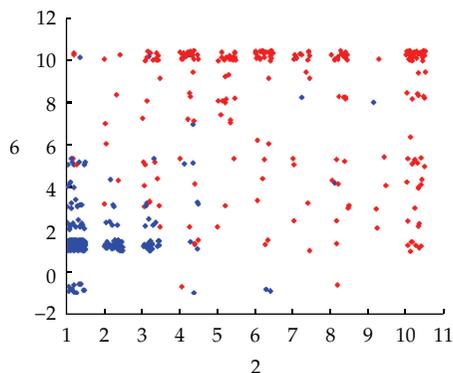


Figure 9: Wisconsin breast cancer data with perturbation. Attributes 2 and 6.

5. Example and Discussion

The Wisconsin breast cancer data set consists of 699 samples with 9 attributes. The samples are classified as two categories, the malignant and the benign. The nine attributes are Clump-Thickness, Cell-Size-Uniformity, Cell-Shape-Uniformity, Marginal-Adhesion, Single-Epi-Cell-Size, Bare-Nuclei, Bland-Chromatin, Normal-Nucleoli, and Mitoses. The data is shown in Figure 9 with an added noise of $[0, 0.5]$.

Now we choose attributes 2 and 6 and accumulate the data samples and use the grid technique proposed in this paper. Then the data set is shown in the following matrix:

$$M_{\text{data}} = \begin{bmatrix} 336 & 13 & 8 & 4 & 9 & 0 & 0 & 0 & 0 & 3 \\ 31 & 4 & 3 & 2 & 0 & 1 & 1 & 1 & 0 & 2 \\ 19 & 5 & 6 & 2 & 4 & 0 & 1 & 1 & 1 & 13 \\ 5 & 1 & 3 & 1 & 3 & 0 & 3 & 2 & 1 & 19 \\ 0 & 1 & 1 & 0 & 1 & 0 & 3 & 5 & 2 & 17 \\ 2 & 0 & 1 & 1 & 1 & 2 & 0 & 0 & 1 & 17 \\ 1 & 1 & 1 & 0 & 2 & 0 & 0 & 1 & 2 & 11 \\ 1 & 1 & 3 & 3 & 2 & 0 & 0 & 4 & 0 & 14 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 7 & 3 & 1 & 5 & 7 & 1 & 0 & 6 & 2 & 35 \end{bmatrix}. \quad (5.1)$$

Next we construct a simplicial complex K as shown in Figure 10 in R^2 where each vertex corresponds to the data points as in (5.1).

We now define $L = 9$. Then $2^L = 512$ and $N = 100 - 25 = 75$. Now we choose $\delta_0 = 1$. Then the final cluster is shown in Figure 11. We find five clusters and 31 outliers.

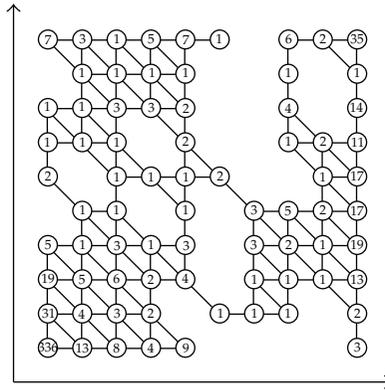


Figure 10: Simplices of the Wisconsin breast cancer data.

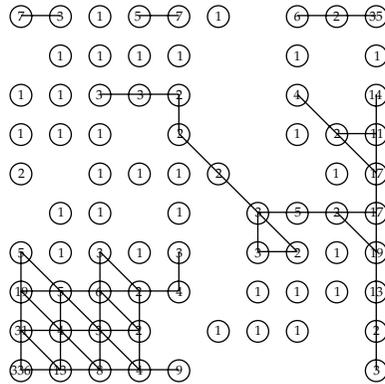


Figure 11: Cluster result of the Wisconsin breast cancer data.

Next we choose attributes 1 and 2 and then the data matrix is

$$M_{\text{data}} = \begin{bmatrix} 131 & 8 & 3 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 41 & 1 & 5 & 0 & 2 & 0 & 1 & 0 & 0 & 0 \\ 80 & 10 & 8 & 3 & 1 & 2 & 1 & 0 & 0 & 3 \\ 56 & 7 & 3 & 3 & 2 & 2 & 1 & 4 & 0 & 2 \\ 63 & 11 & 17 & 8 & 4 & 4 & 5 & 5 & 0 & 13 \\ 10 & 2 & 4 & 0 & 2 & 3 & 0 & 3 & 1 & 9 \\ 1 & 2 & 2 & 4 & 5 & 4 & 0 & 4 & 1 & 0 \\ 0 & 3 & 4 & 7 & 2 & 4 & 7 & 4 & 1 & 14 \\ 1 & 0 & 0 & 1 & 3 & 1 & 1 & 2 & 1 & 4 \\ 1 & 1 & 6 & 13 & 8 & 6 & 3 & 7 & 2 & 22 \end{bmatrix}. \tag{5.2}$$

Now we analyze the effect of the parameter δ_0 . First we choose $\delta_0 = 0$. Then the clustering result is one cluster with all data. Then the ill clustered points is 699. Next let $\delta_0 = 1$. Then we get two clusters and some outliers. However, in this case the red data are all ill-clustered and hence the error rate is 241. Now choose $\delta_0 = 2$. In this case we find four clusters. The number of outliers are 46 while the ill-clustered points are $65 + 74 = 139$. Then

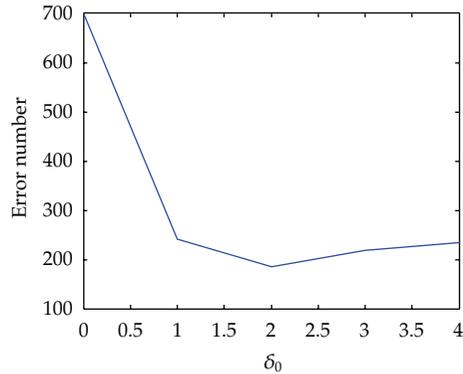


Figure 12: Error number with respect to the parameter δ_0 .

the total number is 185. Next choose $\delta_0 = 3$. This time we find 6 clusters. Number of outliers is 87. Number of ill-clustered points is 91. Number of other error points is 40 hence the total number is 218. Again we choose $\delta_0 = 4$. In this time we obtain 4 clusters with a large outlier (outlier number: 157, error clusters: 65, other errors: 13, total number: 235). The error line is shown in Figure 12.

As a result we see that the best parameter is $\delta_0 = 2$.

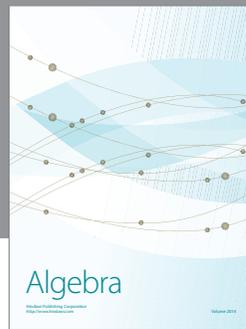
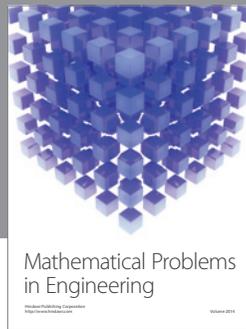
Acknowledgments

This research is supported by the Natural Science Foundation of China (no. 61170038,71071090,60873058), the Natural Science Foundation of Shandong Province (no. ZR2011FM001), and the Shandong Soft Science Major Project (No. 2010RKMA2005).

References

- [1] G. Păun, G. Rozenberg, and A. Salomaa, *Membrane Computing*, Oxford University Press, New York, NY, USA, 2010.
- [2] L. M. Adleman, "Molecular computation of solutions to combinatorial problems," *Science*, vol. 266, no. 5187, pp. 1021–1024, 1994.
- [3] R. J. Lipton, "DNA solution of hard computational problems," *Science*, vol. 268, no. 5210, pp. 542–545, 1995.
- [4] G. Păun, G. Rozenberg, and A. Salomaa, *DNA Computing, New Computing Paradigms*, Springer, Berlin, Germany, 2010.
- [5] T.-O. Ishdorj, A. Leporati, L. Pan, X. Zeng, and X. Zhang, "Deterministic solutions to QSAT and Q3SAT by spiking neural P systems with pre-computed resources," *Theoretical Computer Science*, vol. 411, no. 25, pp. 2345–2358, 2010.
- [6] P. Linqiang and A. Alhazov, "Solving HPP and SAT by P systems with active membranes and separation rules," *Acta Informatica*, vol. 43, no. 2, pp. 131–145, 2006.
- [7] P. Linqiang and C. Martín-Vide, "Solving multidimensional 0-1 knapsack problem by P systems with input and active membranes," *Journal of Parallel and Distributed Computing*, vol. 65, no. 12, pp. 1578–1584, 2005.
- [8] T.-O. Ishdorj, A. Leporati, P. Linqiang, and J. Wang, "Solving NP-Complete problems by spiking neural p systems with budding rules," in *Proceedings of the 10th International Conference on Membrane Computing*, pp. 335–353, 2009.
- [9] P. Linqiang, X. Zeng, X. Zhang, and Y. Jiang, "Spiking neural P systems with weighted synapses," *Neural Processing Letters*, vol. 35, no. 1, pp. 13–27, 2012.

- [10] G. Păun, "A quick introduction to membrane computing," *The Journal of Logic and Algebraic Programming*, vol. 79, no. 6, pp. 291–294, 2010.
- [11] F. Robin, "Morse theory for cell complexes," *Advances in Mathematics*, vol. 134, no. 1, pp. 90–145, 1998.
- [12] F. Robin, "Users guide to discrete Morse theory," *Séminaire Lotharingien de Combinatoire*, vol. 48, article B48c, pp. 1–35, 2002.
- [13] J. Han and M. Kamber, *Data Mining, Concepts and Techniques*, Morgan Kaufmann Publishers, Higher Education Press, Beijing, China, 2002.
- [14] L. Zhang and H. Jiang, "Impulsive cluster anticonsensus of discrete multiagent linear dynamic systems," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 857561, 11 pages, 2012.
- [15] J. Feng, J. Wang, C. Xu, and F. Austin, "Cluster synchronization of nonlinearly coupled complex networks via pinning control," *Discrete Dynamics in Nature and Society*, vol. 2011, Article ID 262349, 23 pages, 2011.
- [16] W. Zou, Y. Zhu, H. Chen, and X. Sui, "A clustering approach using cooperative artificial bee colony algorithm," *Discrete Dynamics in Nature and Society*, vol. 2010, Article ID 459796, 16 pages, 2010.
- [17] X. Liu, L. Xiang, X. Wang et al., "Spatial cluster analysis by the adleman-lipton DNA computing model and flexible grids," *Discrete Dynamics in Nature and Society*, vol. 2012, Article ID 894207, 32 pages, 2012.
- [18] C. Mónica, M. Angels Colomer, M. J. Pérez-Jiménez, and A. Zaragoza, "Hierarchical clustering with membrane computing," *Computing and Informatics*, vol. 27, no. 3+, pp. 497–513, 2008.
- [19] R. Ceterchi and C. Martín-Vide, "P systems with communication for static sorting," GRLMC Report 26, Rovira I Virgili University, 2003, M. Cavaliere, C. Martín-Vide, and Gh. Paun, eds.
- [20] R. B. Abu Bakar, J. Watada, and W. Pedrycz, "DNA approach to solve clustering problem based on a mutual order," *BioSystems*, vol. 91, pp. 1–12, 2008.
- [21] V. Natarajan and H. Edelsbrunner, "Simplification of three-dimensional density maps," *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, pp. 587–597, 2004.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

