Asian Journal of Chemistry Vol. 19, No. 3 (2007), 2381-2388

A Note on the Topological Sense of Chemical Sets

ÖMER GÖK* and MESUT KARAHAN†
Department of Mathematics, Faculty of Arts and Sciences, Yildiz Technical University
Davutpasa Campus, 34210 Esenler-Istanbul, Turkey
E-mail: gok@yildiz.edu.tr

The present studies suggested the way to evaluate some topological properties, such as; closure, derived set, boundary, interior, exterior, semi-openness, semi-closednes, semi-regularity, θ-semi closure, θ-semi openness, regular opennes of subsets of chemical set interested in the chemical set chosen at the beginning as object in the work of Restrepo *et.al*.

Key Words: Topology, Semi-open sets, Semi-regular sets, Dendrograms, Mathematical chemistry.

INTRODUCTION

It is known that there are many chemical systems determined by the similarity relationships among their elements, for example; groups of chemical elements, alkanes, acids, ketones, *etc*. When we ask how to quantify such a relationship and if they are related, we say that two elements of a set of chemical interest are very similar. It is possible to define each element as a point in a mathematical space and it is evaluated its relationship with others by means of a distance function[1-3]. In this method each element is determined by using several features of itself. Also, the number of these properties defines the dimension of the space where we consider that element as a point. For this methodology we refer to cluster analysis[1,2,4]. As independent of the dimension of the space, one way to illustrate such clusters is a two-dimensional graphic representation by means of dendrogram[5]. In work of many researchers[5-8], it is to interpret a dendrogram and its clusters as a map of neighbourhoods of the elements, and extracting from these clusters such similarity neighbourhoods. By using a neighbourhood of every element of the set, we can apply topology in charge of studying neigbourhood relation[5-8]. Here, it is possible to determine, topologies on the set and to study some topological properties of itself: closure, derived set, boundary (frontier), interior, exterior, semi-open, semi- closed, semi-closure, semi-regular, regular open and regular closed. For a recent result, we refer the reported work[5-8]. Taking a topological approach of a dendro-

†Department of Chemistry, Faculty of Arts and Sciences, Yildiz Technical University, Davutpasa Campus, 34210 Esenler-Istanbul, Turkey, E-mail: mesut.karahan@gmail.com

gram (complete binary tree), it is possible to find out well-known relationships. As an application of this methodology it can be done a mathematic interpretation of a dendrogram where there are no arbitrarities in the interpretation of clusters[9,10].

In this paper, we improve the work G. Restrepo *et. al.*[10], in the new-topological properties of a set.

Preliminaries

Here, it is introduced well-known results from[10] called methodology. If we have a chemical set Q of m elements x_i, where every one is defined as a vector $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ of its n properties, then we can apply cluster analysis to this set with the purpose of knowing the similarity relationships among x_i's. First, we build up a matrix of elements ($m \times n$) and calculate by means of a similarity function[1-3,5-8] (frequently a metric[9]) the similarity among all the elements. Thus, we build up a new matrix ($m \times m$) called similarity matrix[1,4,11] and using a grouping methodology, we obtain clusters of elements. These clusters are represented in a dendrogram, which is independent of the dimension n of the space of work because it is always two-dimensional. In Fig. 1, it is an example of one of them.

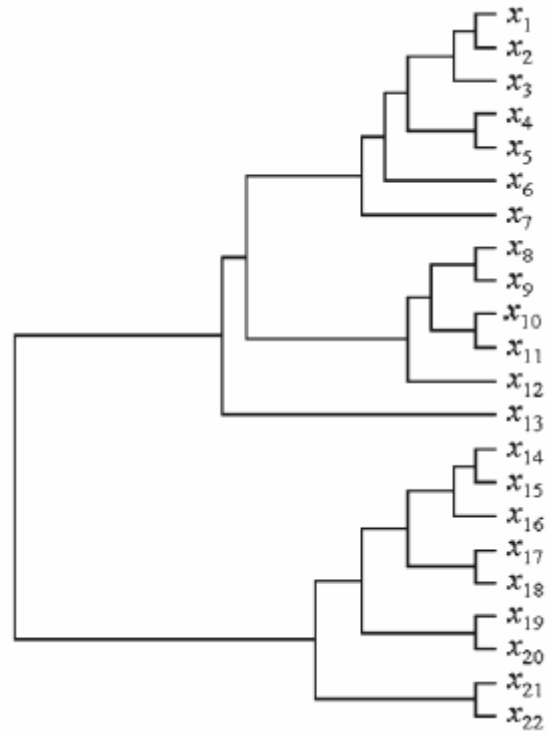

Fig. 1. A Dendrogram

It is seen from Fig. 1 that the neighbourhood of x_1 is itself and x_2, or if not strict, x_3 belongs to the neighbourhood of x_1. an element belongs to the neighbourhood of another element if they both belong to the same "branch" in the dendrogram[5,8,12]. In a recent work[6], it is shown a methodology to define these branches as subtrees and introduced a mathematical method to define these subtrees as subgraphs[5].

Codes on the dendrogram

We associate a code made of 0s and 1s to every element on the dendrogram. In figure 3 the dendrogram of figure 1 is codified.

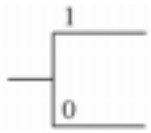

Fig. 2. Codification system

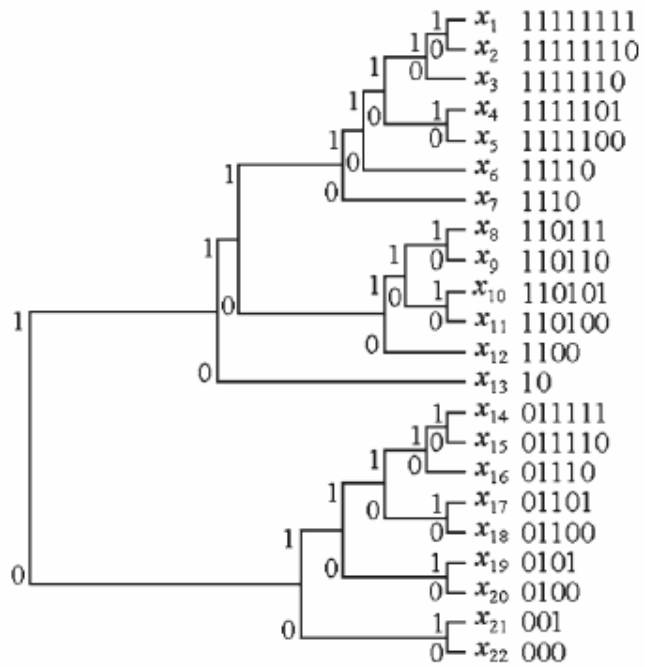

Fig. 3. Codes on dendogram of Fig. 1

It is important to remark that the identity of every element is characterized by its code because there is only one code for each element, although the system of codification changes. It means, if we put 0 "above" and 1 "below," the code of every element changes but its identity remains the same. In spite of its changes of code, there is only one identity for every element.

Now, with these codes we can talk about neighbourhoods in terms of codes. In other words, we can put the intuitive idea of neighbourhood as a branch of the dendrogram in a numerical way. In the following we develop this idea defining subtree in terms of codes.

A subset A of Q is called subtree if there is a code $\alpha_1\alpha_2...\alpha_k$ such that:

1. If $x \in A$, then the former components of code of x coincide with $\alpha_1\alpha_2...\alpha_k$.
2. If for each $y \in Q$ the former components of code of y coincide with $\alpha_1\alpha_2...\alpha_k$, then $y \in A$.

A subtree is the set $\{y \in Q | y$ starts with the code $\alpha_1\alpha_2...\alpha_k\}$.

In Fig. 3, the subset $R = \{x_8,x_9,x_{10},x_{11}\}$ is a subtree since all its elements start from the code 1101.

The subset NR = $\{x_1, x_2, x_4, x_5\}$ is not a subtree.

An n-subtree is a subtree of cardinality less than or equal to n. In other words, an n-subtree has at most n elements. The subtree R of example 1 is an example of 4-subtree or 5-subtree. A maximal n-subtree is an n-subtreesuch that there is no other n-subtree containing it. The subtree R is a maximal 4-subtree.

Proposition 1: B_n is a partition of Q, where $B_n = \{B|B$ is a maximal n-subtree$\}$.

Lemma 2: Every partition defined over a subset is basis for a topology.

B_n produce a topology on Q by means of arbitrary unions among its elements. Now, define the topology on Q obtained using basis by Lemma 2.

Let $\tau_n = \{\cup_{B\in F}B \mid F \subseteq B_n\}$ a topology on the set Q.

Proposition 3: The pain (Q, τ_n) is a topological space[10].

On the chemical meaning of topological properties

In a work[5-8] they showed that several intuitive chemical ideas can be explained according to our methodology. For example, they found that the mathematical boundary of metals and non-metals is the same set of chemical elements, that is semimetals[5-8]. This result shows that the ancient concept of semimetal as an element whose properties are not from metals nor from nonmetals has a mathematical explanation taking advantage of known properties of chemical elements. On the other hand, they showed[8] that, taking advantage of results from Molecular Quantum Similarity[13], the intuitive classification of steroids according to chemical knowledge on structure and reactivity gives a disjoint set which indicates that this classification is correct. They apply the same methodology to sets of amino acids[14] and benzimidazoles[15] and they found[8] some chemical species belonging to more than one set; something like happens with semimetals regarding to metals and non-metals[5-8]. These results indicate that there are some substances sharing their properties with substances of other sets commonly

considered different. These results arose from topological properties of sets *Q* of chemical interest. Let us take the following example:

Example 4: Let $Q = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z\}$. See Fig. 4, if $n = 4$, then we research 4-maximal subtrees on the dendrogram. Then,

$B_4 = \{\{a, b, d\}, \{f, m\}, \{c, g, l\}, \{e, n, p\}, \{t, i\}, \{u, h, o\}, \{v, w\}, \{x, y, q, z\}, \{s, j, k\}, \{r\}\}$ is a basis for a topology. Let us call *A*, which is: $A = \{a, b, c, d, e, f, g\} \subset Q$. We are interested in the topological properties of *A*.

Closure of a subset

We denote it by $\hat{A}$ or Cl(*A*). Here, $\hat{A} = \{a, b, c, d, e, f, g, l, m, n, p\}$.

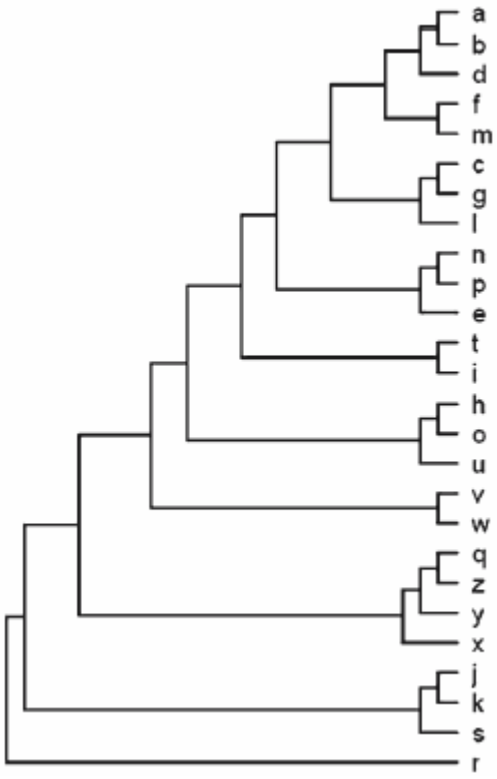

Fig. 4. A particular dendrogram of 26 elements

Derived set of a subset

We denote it by *A*'. Here $A' = \{a, b, c, d, g, l, m, n, p\}$.

Boundary of a subset

We denote it by *b(A)*. $b(A) = \{c, e, f, g, n, p, m, l\}$

Interior of a subset

We denote it by Int(*A*). $\text{Int}(A) = \{a, b, d\}$

Exterior of a subset

We denote it by Ext(A). Ext(A) = {h, i, j, k, o, q, r, s, t, u, v,w, x, y, z}.

A semi-openness of a subset

We denote it by sInt(A). Since $A \not\subset$ sInt(A), A is not a semi-open set.

A semi-closure of a subset

We denote it by sCl(A). A is not a semi-closed set, because its complement is not a semi-open set.

A θ-semi-closure of a subset

We denote it by θ-sCl(A). θ-sCl(A) ≠ A.

Regular openness of a subset

We denote it by RO(A). RO(A) = Int(Cl(A)) = A

$\hat{A}$ or Cl(A) = {a, b, c, d, e, f, g, l,m, n, p}.

RO(A) ≠ A.

Regular closedness of a subset

We donote it by RCl(A).

RCl(A) ≠ A.

Let X be a non-empty set and P a collection of subsets of X. P is called a partition of X iff:

1. $X = \cup_{B\in P}B$
2. If B_1 and $B_2 \in P$, then $B_1 \cap B_2 = \emptyset$

Let X be a non-empty set and τ a collection of subsets of X such that:

1. $X \in \tau$
2. $\emptyset \in \tau$
3. If $O_1, \ldots, O_n \in \tau$, then $\cap^{n}_{j=l} Oj \in \tau$
4. If $\alpha \in I$, $O_\alpha \in \tau$, then $\cap_{\alpha\in I} O_\alpha \in \tau$.

Thus, τ is a topology, the couple (X, τ) is called a topological space and the elements of τ are called open sets.

Let B be a collection of subsets of a non-empty set X, such that:

1. X = $\cup_{B\in B}B$
2. If B_1,$B_2 \in$ B, then $B_1 \cap B_2$ is the union of elements of B, then B is called a basis for the topology τ,

where τ = {$\cup_{B\in F}B \mid F \subseteq$ B}.

Some topological properties are the following:

Let $A \subset X$ and $x \in X$; x is said to be a closure point of A iff for every $O \in \tau$, such that $x \in O$, then

$O \cap A \neq \emptyset$.

Let A $\subset X$; the closure of A is defined as: Cl(A) = $\hat{A}$ = {$x \in X \mid x$ is closure point of A}.

Let A $\subset X$ and $x \in X$; it is said that x is an accumulation point of A iff for every $O \in \tau$, such that

$x \in O$, then $(O - \{x\}) \cap A \neq \emptyset$.

Let $A \subset X$; the derived set of A is defined as: A'= $\{x \in X \mid x$ is accumulation point of $A\}$.
Let $A \subset X$ and $x \in X$; it is said that x is a boundary point of A if for every $O \in \tau$, such that $x \in O$, then
$O \cap A \neq \emptyset$ and $O \cap (X - A) \neq \emptyset$.
Let $A \subset X$; the boundary of A is defined as:
$b(A) = \{x \in X \mid$ x is boundary point of $A\}$.
Let $A \subset X$ and $x \in X$; it is said that x is an interior point of A if for every $O \in \tau$, such that $x \in O$, then
$O \cap (X - A) = \emptyset$.
Let $A \subset X$; the interior of A is defined as:
Int(A) = $\{x \in X \mid x$ is interior point of $A\}$.
Let $A \subset X$ and $x \in X$; it is said that x is an exterior point of A iff for every $O \in \tau$, such that $x \in O$, then
$O \cap A = \emptyset$.
Let $A \subset X$; the *exterior* of A is defined as:
Ext(A) = $\{x \in X \mid x$ is exterior point of $A\}$.

Let X be a topological space and let S be a set in X. S is called semi-open if S $\subset$ Cl(Int(S))[16,17]. The complement of a semi-open set is called semi-closed[18]. The intersection of all semi-closed sets of X containing S is called the semi-closure of S[17,18] and denote it by sCl(S).

The family of all semi-open (respectively open) subsets of X is denoted by $SO(X)$ (respectively $O(X)$). Set $SO(X, x)=\{U \mid x \in U \in SO(X)\}$. A subset S of X is said to be semi-regular if it is semi open and semiclosed[19].

The family of all semi-regular subsets of X is denoted by SR(X). A point $x \in X$ is said to be the θ- semi-cluster point of S $\subset X$ if Cl(U) $\cap$ S $\neq \emptyset$ for every $U \in SO(X, x)$[20]. The set of all θ-semi-cluster points of S is called the θ-semi-closure of S and is denoted by θ-sCl(S). If S=θ-sCl(S), then S is called θ-semi-closed. The complement of a θ-semi-closed set is called θ-semi-open. A subset S of X is called regular open if S= Int(Cl(S)). Its complement is called regular closed, *i.e.* S= Cl(Int(S)).

Conclusion

In this work, chemical objects were studied by means of discrete mathematics and we add a development of a mathematical methodology to study sets of chemical objects Q based on a classification of elements in Q to work G. Restrepo *et al.*[10]. It is shown a way[6] to represent every element of the tree as a code and defined neighbourhoods on the tree representing neighbourhoods on the space of all elements of Q. We described a procedure to calculate some topological properties of subsets of Q, such as closure, derived set, boundary, interior, exterior, semi-openness, semi-closedness, semi-regularity, θ-semi closure, θ-semi openness, regular openness and so on. It is shown the chemical meaning of such properties if

each element of *Q* has been defined according to its features or properties. As applied this methodology to chemical systems, it can be applied the same procedure to other systems such as physical, biological. For this, we must have a set of discrete elements that can be classified and shown as a tree.

REFERENCES

1. R.R. Sokal and P.H.A. Sneath, Principles of Numerical Taxonomy, Freeman, San Francisco (1963).
2. J.M. Barnard and G.M. Downs, *J. Chem. Inf. Comput. Sci.*, **32**, 644 (1992).
3. P. Willet, *J. Chem. Inf. Comput. Sci.*, **38**, 983 (1998).
4. M. Otto, Chemometrics, Statistics and Computer Application in Analytical Chemistry, Wiley-VCH, Weinheim (1999).
5. G. Restrepo, H. Mesa, E.J. Llanos and J.L. Villaveces, in eds.: R.B. King and D. Rouvray, The Mathematics of the Periodic Table, Nova, New York, Ch. 5, (2004).
6. G. Restrepo, H. Mesa, E.J. Llanos and J.L. Villaveces, *J. Chem. Inf. Comput. Sci.*, **44**, 68 (2004).
7. G. Restrepo, H. Mesa, E.J. Llanos and J.L. Villaveces, Proceedings of the Third Joint Sheffield Conference on Chemoinformatics, The University of Sheffield, Shefiield (2004).
8. G. Restrepo and J.L. Villaveces, Proceedings of the Nineteenth International Course & Conference on the Interfaces among Mathematics, Chemistry & Computer Sciences, Inter-University Center, Dubrovnik, (2004).
9. B. Mendelson, Introduction to Topology, Dover, New York (1990).
10. G. Restrepo, H. Mesa and J. L. Villaveces, *J. Math. Chem.*, **39**, 363 (2006).
11. J.V. Crisci, Introducci´on a la teor´ia y pr´actica de la taxonom´ia num´erica, OEA, Washington (1983).
12. H. Mesa and G. Restrepo, Proceedings of the Primer Encuentro Nacional de Qu´imicos Te´oricos, Universidad de Pamplona, Pamplona (2004).
13. P. Bultinck and R. Carbo-Dorca, *J. Chem. Inf. Comput. Sci.*, **43**, 170 (2003).
14. C. Cardenas, M. Obregon, E.J. Llanos, E. Machado, H.J. Bohorquez, J.L. Villaveces and M.E. Patarroyo, *Comp. Chem.*, **C26**, 667 (2002).
15. M. Nino, E.E. Daza and M. Tello, *J. Chem. Inf. Comput. Sci.*, **41**, 495 (2001).
16. N. Levine, *Am. Math. Monthly*, **70**, 36 (1963).
17. S. Jafari and T. Noiri, *Demonstration Math.*, **34**, 187 (2001).
18. S.G. Crossley and S.K. Hildebrand, *Texas J. Sci.*, **22**, 99 (1971).
19. G. Di Maio and T. Noiri, *Indian J. Pure Appl. Math.*, **18**, 226 (1987).
20. J.E. Joseph and M.H. Kwack, *Proc. Am. Math. Soc.*, **80**, 341 (1980).

(*Received*: 11 July 2006; *Accepted*: 4 November 2006) AJC-5258