Theoret. Chim. Acta (Berl.) 55, 55–75 (1980)

THEORETICA CHIMICA ACTA

The Structures of Molecular Topological Spaces

Richard E. Merrifield and Howard E. Simmons

Central Research & Development Department, E. I. du Pont de Nemours & Company, Experimental Station, Wilmington, Delaware 19898, USA

The concepts of general topology are employed to derive a mathematical description of the structures of molecules. It is shown that a topological space on a finite set of points induces a unique graph and that as a consequence there is a uniquely determined topological space associated with every alternant molecule. This space is shown to be identical to the quotient space which results from partitioning the region of real space occupied by a molecule into atomic subregions. The molecular topological space is connected if and only if the molecule is connected and the only molecules having equivalent topological spaces are stereoisomers. Nonalternants are topologically distinguished from alternants by the fact that their graphs are not derivable from a topological space as are those of alternants. A set of graph-theoretical techniques for analyzing the combinatorial structure of finite topologies is developed. The cardinality of the molecular topology is found to be a measure of molecular complexity and the cardinalities of the subspace topologies associated with the bonds of the molecule are accurate measures of relative bond strength. Several empirical correlations between physical properties of molecules and topological measures are found.

Key words: Topological spaces, molecular ~.

1. Introduction

To what extent are the properties of a molecule a consequence of the way in which its parts are linked together and to what extent do they depend on the metric properties of bond lengths and angles and on the detailed dynamics of electrons and nuclei? These aspects are all inextricably bound up together in a correct quantum mechanical description of the structure of a molecule. In this paper we present a new approach to this question using the mathematical tools of topology, a subject

0040-5744/80/0055/0055/$04.20

specifically designed to address those aspects of structure which are intrinsically nonmetric.

Some aspects of this question have been dealt with in the past by a variety of approaches. Many of these have involved devising various empirical indices of molecular branching or complexity and seeking correlations of these with observed molecular properties [1]. More mathematically sophisticated treatments have been based on graph theory [2], although, with the notable exception of the work of Hosoya [3], most of these have been essentially restatements of Hückel molecular orbital theory. We believe that topology is a peculiarly appropriate tool for attacking this problem and have shown previously [4] that it is indeed possible to extract meaningful structural information from such an approach.

2. Mathematical Preliminaries

We present here the basic ideas from topology which will be employed in this paper. Most of these are described in detail in introductory topology texts [5], however some of the notions that are peculiar to spaces with a finite number of points are not generally found in textbooks.

The basic concept through which topology expresses the idea of structure is that of a *topological space*. This consists of a structureless set of points X together with a *topology* $\mathscr{T}$, which is a collection of subsets of X (called *open sets*). As far as topology is concerned the points of X are undefined primitives; it is in the applications that they are given a specific physical interpretation as, e.g., atoms or orbitals. The open sets impart a structure to X in the sense that points belonging to the same open set are regarded as being near to each other (any open set containing the point x is called a *neighborhood* of x) even though no quantitative measure of distance is introduced. The open sets are required to conform to the following restrictions:

1. The empty set $\varnothing$ and X itself are open.
2. Arbitrary unions of open sets are open.
3. Finite intersections of open sets are open.

Even with these restrictions, there is still a very large number of possible topologies for a given set; it is largely up to one's ingenuity to select those which reflect those aspects of structure regarded as significant.

A *closed set* is a subset of X whose complement, i.e., the set of elements of X which do not belong to the given set, is an open set. In spaces with a finite number of points the collection of closed sets also forms a topology $\mathscr{T}^*$, called the *cotopology*.

Rather than dealing with the entire collection of open sets it is usually more convenient to work with a smaller collection, called a *basis* for a topology, defined by the requirement that the collection of all unions of basis elements (including the empty union, $\varnothing$) is equal to $\mathscr{T}$. For finite spaces it is useful further to consider the *irreducible basis* for a topology, defined as a basis no element of which can be expressed as a union of other basis elements. In contrast to a basis, the irreducible basis of a topology is unique. A *subbasis* for a topology is a collection of sets, all of whose intersections taken together form a basis. Any collection of sets whose union is X is a subbasis for some topology.

For an arbitrary subset $A \subset X$, the *closure* $\bar{A}$ is the smallest closed set such that $A \subset \bar{A}$. Similarly, in a finite space the smallest open set $\hat{A}$ such that $A \subset \hat{A}$ is called the *coclosure* since it is the closure of A in the cotopology. The relation of these concepts to that of a basis is given by the following [6]:

1. The irreducible basis of a finite topology is equal to the collection of coclosures of the one-point sets.
2. The irreducible basis of a cotopology is equal to the collection of closures of the one-point sets.

Thus, for the topology, there is an irreducible basis element $B_p = \widehat{\{p\}}$ associated with each point $p \in X$ and, similarly, for the cotopology the irreducible basis elements are $B_p^* = \overline{\{p\}}$.

A *subspace* of a topological space $(X, \mathscr{T})$ is a subset $Y \subset X$ together with a topology $\mathscr{T}_Y$ defined by

$$\mathscr{T}_Y = \{Y \cap O_i | O_i \in \mathscr{T}\}.$$

In the present context the notion of a subspace arises in connection with the problem of constructing the collection of open sets of $(X, \mathscr{T})$ which contain a given $A \subset X$ as a subset. This collection of sets, which is given by $\{\hat{A} \cup O_i \mid O_i \in \mathscr{T}\}$, is in fact isomorphic to the topology of the subspace $X - \hat{A}$. The correspondence between these two collections of sets is given by the set-theoretic identity

$$(X - \hat{A}) \cap O_i = (\hat{A} \cup O_i) - \hat{A}.$$

A topological space $(X, \mathscr{T})$ is said to be *connected* if X cannot be written as a union of disjoint open sets. For a space which is not connected, the connected open sets into which it can be partitioned are its *components*. A space is said to be T_0 if it satisfies the axiom: Given any pair of distinct points, there exists an open set which contains one of the points but not the other. (Stronger separation axioms, e.g., the Hausdorff axiom, are of little interest in finite spaces as they are satisfied only by the *discrete* topology, i.e., the collection of all subsets of X.)

Many aspects of finite topologies can be given convenient graph-theoretic representations. If the points of the space are represented by the vertices of a graph and there is an edge joining vertices p and q whenever either $B_p \subset B_q$ or $B_p \supset B_q$, then the result is the *comparability graph* of the irreducible basis. If $B_p \supset B_q$, then there are no open sets which contain p but not q and hence no closed sets which contain q but not p. This implies $B_p^* \subset B_q^*$ so that the inclusion relations among the cotopology basis elements are the inverse of those for the topology and, in particular the comparability graphs of topology and cotopology are isomorphic. The direction of the inclusion relations can be indicated by a directed graph (digraph) in which $B_p \supset B_q$ is represented by an edge from p to q. The digraph for the cotopology is then the converse of that for the topology, i.e., it is the latter digraph with the orientation of each edge reversed. Since inclusion is transitive, these digraphs must be transitive; it has also been shown that every transitive digraph corresponds to a topology [7]. Finally, a comparability graph has the same number of components in the graph-theoretical sense as its topological space has topological components since a component of X corresponds to a set of vertices none of which are comparable to the vertices of any other component. Thus, in particular, a connected space corresponds to a connected comparability graph.

3. A Topology for Alternant Molecular Graphs

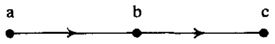
To illustrate the application of topological concepts to molecular structure, consider the carbon skeleton of a saturated or a conjugated, unsaturated hydrocarbon represented as a graph G, the vertices of which represent the carbon atoms in the former case and either carbon atoms or ***p***-orbitals in the latter and the edges of which represent the atomic adjacency relations. We shall regard the vertices as the points of X and we seek to construct a topology $\mathscr{T}$ which reflects the molecular structure. One approach is to start from a collection of sets which express the atomic connectivity. In the "bond topology" [4] this was accomplished by taking the collection of 2-sets representing the edges of G as a subbasis for $\mathscr{T}$. This procedure guarantees that all bonds are open sets and the resulting topology, with a suitably defined bond "weight" function, gives an excellent account of bond strength patterns for small π-systems. In spite of its successes, the bond topology has some unsatisfactory features:

1. Connected molecules do not always yield connected spaces.
2. There is a high incidence of homeomorphism, that is, distinct molecules that are topologically indistinguishable.
3. In a given molecule there are at most seven topologically distinct bonds so that subtle variations in bond strength in large molecules are lost.

These characteristics are all, in one way or another, consequences of the fact that the bond topology yields highly disconnected spaces which are made up of only three different kinds of components containing 1, 2 or 3 atoms each. It would be highly desirable to be able to describe a molecule in terms of a connected topological space.

The discussion in Sect. 2 suggests an alternate way of topologizing a molecular graph which will achieve this objective. Specifically, since there is a unique basis comparability graph associated with every finite topology, we consider the converse problem and ask if the graph of any molecule can be regarded as the comparability graph of some topology, which we shall call the *graph topology* for that molecule. Unfortunately, not all molecular graphs can be so regarded, but only those whose edges can be oriented to yield a transitive digraph.

Not every graph can be transitively oriented. For example, the odd cycles C_5, C_7, . . . cannot be. Any orientation of their edges must contain somewhere the configuration



but this is not transitive since the edge (a, c) is absent. This suggests that transitively orientable graphs in general are characterized by a plethora of 3-membered rings.

Consider the subclass of transitively orientable graphs which contain no 3-membered rings. When oriented, such a graph must not contain the above forbidden configuration. The only possibility is that for every vertex either the in-degree or the out-degree must be zero. Thus it must be possible to partition the vertices of the graph into two sets such that no two vertices in the same set are adjacent, i.e., the graph is bipartite or, in the case of π-electron systems, the molecule is alternant. Furthermore, the graph has exactly two transitive orientations and these are converses of each other.

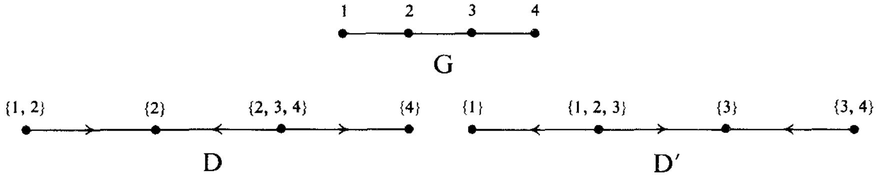
Alternants are thus topologically distinguished from nonalternants (or, at least, those nonalternants containing cycles of length greater than three[1]) by the fact that their graphs are derivable as the comparability graph of a topological space.

To the two transitive orientations, D, D' of a bipartite graph G, there correspond a unique topology-cotopology pair for which D and D' express the inclusion relations among the irreducible basis elements. Recalling from Sect. 2 that an edge (p, q) in D means that $B_p \supset B_q$ the basis elements of the topology induced by D are given by

$$B_p \equiv \widehat{\{p\}} = \{p\} \cup \{q \mid q \text{ adjacent from } p\}.$$

Thus vertices of zero out-degree are open singletons while vertices of zero in-degree are closed singletons.

These considerations can be illustrated by the following example:



where the basis sets induced by D and D' are indicated above the corresponding atoms. The topology corresponding to D is:

$$\mathscr{T} = \{\varnothing, \{2\}, \{4\}, \{1, 2\}, \{2, 4\}, \{1, 2, 4\}, \{2, 3, 4\}, \{1, 2, 3, 4\}\}$$

and the topology induced by D', which is the cotopology of $\mathscr{T}$, is

$$\mathscr{T}^* = \{\varnothing, \{1\}, \{3\}, \{1, 3\}, \{3, 4\}, \{1, 2, 3\}, \{1, 3, 4\}, \{1, 2, 3, 4\}\}.$$

(Since $\mathscr{T}$ and $\mathscr{T}^*$ are generated completely symmetrically from G, it is purely a matter of convention which is called the topology and which the cotopology.) Note that while neither $\mathscr{T}$ nor $\mathscr{T}^*$ alone properly reflects the symmetry of G in this example, the totality of open and closed sets does so. (It turns out that, even in

---

[1] Nonalternant structures which contain only 3-membered rings are also transitively orientable and hence are topologizable in the same manner. However a much larger number of topologies is produced, e.g., cyclopropenium has six distinct transitive orientations.

cases where $\mathscr{T}$ and $\mathscr{T}^*$ separately have the correct symmetry, it is still necessary to consider both for a complete description of the structure.)

To gain some insight into the structures of these spaces we need to devise a quantitative measure of the contribution of each bond in the molecule to the space. As such a measure one might take either $n(p, q)$, the total number of open and closed sets containing the bond (p, q), or – as in the bond topology – $c(p, q)$, the sum of their cardinalities. For the current example:

$$n(1,2) = 5, \qquad n(2,3) = 4,$$
$$c(1,2) = 16, \qquad c(2,3) = 14.$$

For comparison the Hückel bond orders are

$$p(1,2) = 0.894, \qquad p(2,3) = 0.447.$$

By either measure, the graph topology gives the correct order, but the magnitude of the variation is too small and not as good as that given by the bond topology. We shall see in further examples that the graph topology in general yields rather small variations among bonds, but it has the great advantage over the bond topology of continuing to yield variations under conditions where the latter gives featureless uniformity.

The two other key features of the graph topology which follow directly from the manner in which it is constructed are:

1. Since a comparability graph and its associated topological space have the same number of components, the space of a molecule is connected if and only if the molecule is connected.
2. A finite topology induces a unique comparability graph; hence distinct molecular graphs have distinct topologies, i.e., the only homeomorphic molecules are stereoisomers.

Thus the graph topology yields a global description of molecular structure which assigns a unique connected space to every alternant molecule. In contrast, the bond topology corresponds to a local description, consisting of many components which reflect primarily the nature of the terminations of the molecule. This can be illustrated with our current example, for which the bond topology is [4]

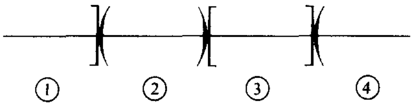
$$\mathscr{T}_{\text{bond}} = \{\varnothing, \{2\}, \{3\}, \{1,2\}, \{2,3\}, \{3,4\}, \{1,2,3\}, \{2,3,4\}, \{1,2,3,4\}\},$$

which corresponds to the digraph

This shows the power of the graph-theoretical approach to finite topologies; it is far from obvious from merely inspecting the collection of open sets whether or not it describes a connected space.

A final topological property of the spaces of the graph topology is that they are T_0 spaces, a property which follows from the fact that their digraphs contain no vertices with both in-degree and out-degree nonzero and hence they contain no directed cycles [7].

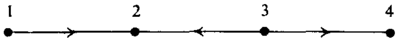
There is a second, quite different way of arriving at the graph topology for molecules which perhaps offers some conceptual advantage over the rather abstract approach just outlined and which also establishes a connection between the finite topological spaces considered in this paper and the continuous spaces which are the usual concern of topology. In this alternative point of view we regard a molecule as a connected region of real space which is partitioned into contiguous, disjoint regions which are identified with the constituent atoms. Our previous example will illustrate the concept. Consider the real line ***R*** partitioned into four regions:



where regions 2 and 4 are open intervals, i.e., they do not contain their boundary points, and regions 1 and 3 are closed intervals. If we regard each region as a point in a new space and construct for each such point the smallest set containing that point which corresponds to an open interval of ***R*** (the open intervals are a basis for the "usual" topology of the real line), we find

①	②	③	④
●	●	●	●
{1, 2}	{2}	{2, 3, 4}	{4}

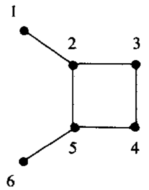
which is the basis of the graph topology of



A space topologized in this manner, by identifying its points with regions of the original space and constructing the basis as illustrated, is called a *quotient space.*

As the example shows, the partition ①](② leads to an edge $\overset{\textcircled{1}}{\bullet}\longrightarrow\overset{\textcircled{2}}{\bullet}$ in the digraph for the quotient space. If real space is partitioned into regions which are either open or closed, the vertices of the digraph will have either zero in-degree or zero out-degree and the topology of the quotient space is indeed the graph topology.

The graph topology can also be regarded as a quotient of three-dimensional real space, $\boldsymbol{R}^3$. The molecule is first enclosed by a closed surface (multiply connected if the molecule contains cycles) which is then partitioned into contiguous open and closed regions corresponding to the atoms. As an illustration, consider the molecule whose graph is



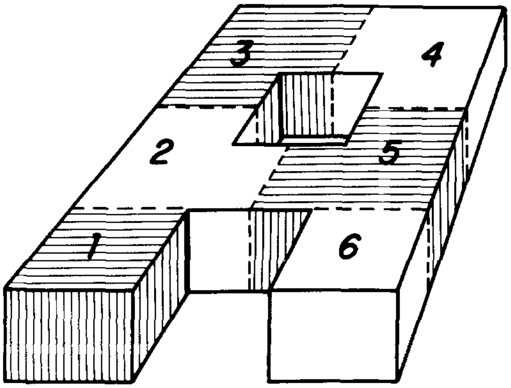
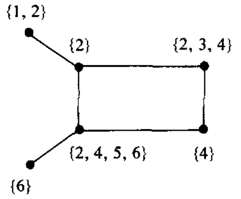
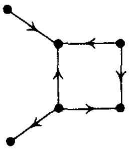


Fig. 1. Partition of real space corresponding to 1,2-dimethylenecyclobutene. The shaded and clear regions denote closed and open volumes corresponding to the individual atoms

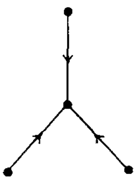
This is represented as a partitioned volume of real space as shown in Fig. 1. As before, we regard each atomic volume as a point and construct for each atom the smallest set containing that atom which corresponds to an open volume of real space. This construction can be shown to yield for the basis of the quotient topology:



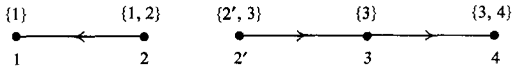
which is the graph topology of



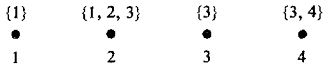
The concept of a quotient space also provides an insight into the relation between the graph topology and the bond topology in the sense that it allows one to build a molecule with the former topology out of pieces having the latter. (The process we are about to describe is known formally as an "attaching map". Rather than give a rigorous discussion we present simply the intuitive ideas involved.) The only three connected molecules in the bond topology have topologies corresponding to the digraphs:



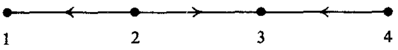
As an example, consider a space consisting of two of these fragments



where the indicated sets are the irreducible basis elements. If we now construct a quotient space in which points 2 and 2′ of the original space are identified with point 2 of the new space we get for the basis of the quotient topology:

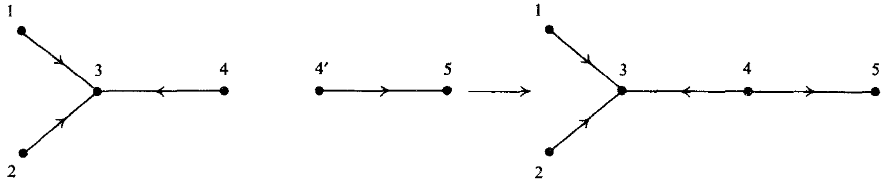


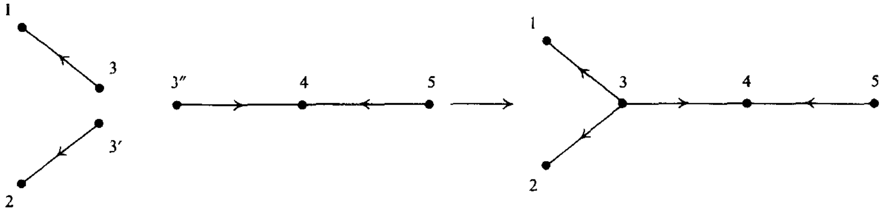
which corresponds to the graph topology of



(This particular type of quotient space formation is called the "pasting" of two spaces.)

A slightly more complicated example of this pasting process is:





Although we choose not to give a formal proof, these examples should make plausible the conclusion that the space of any molecule with the graph topology can be produced in a unique manner by a pasting of the three connected fragments of the bond topology.

4. Combinatorial Properties of the Graph Topology

In spaces of infinitely many points, e.g., real space, topology is usually employed to examine questions of convergence and continuity. In finite topological spaces, on the other hand, the most important properties are combinatorial ones such as the numbers of sets of various kinds. In the bond topology it was possible to give closed-form expressions for the combinatorial properties of a general molecule [4] as a result of the fact that these spaces consist of a limited variety of small components. In the connected spaces of the graph topology such a succinct statement of the results is not possible although the combinatorics are still tractable. In this section we develop the machinery required to analyze the combinatorics of the graph topology.

For an alternant molecular graph G we define the function $\sigma(G)$ to be the number of open sets in the graph topology of G. (Since there is a one-to-one correspondence between open and closed sets, the number of sets in the cotopology is the same, i.e., it does not matter which of the two transitive orientations D, D' of G we consider.) Recall that there is a unique irreducible basis set B_p associated with each vertex p of G and that the topology $\mathcal{T}$ is the collection of all unions of basis sets. However, not all such unions are distinct. Consider an arbitrary union of basis sets:

$$V = B_1 \cup B_2 \cup \cdots.$$

If, for example, $B_1 \subset B_2$ then B_1 can be eliminated from the union without affecting the result. Thus to produce only distinct unions it is required that no two elements entering into the union be comparable. Since comparable basis elements are those corresponding to adjacent vertices of G, distinct open sets are guaranteed if only those unions of basis elements are formed which correspond to sets of vertices of G no two of which are adjacent. (In graph theory such a set of vertices is called a *stable set* of G.) Thus $\sigma(G)$, the cardinality of the graph topology of G, is equal to the number of stable sets of G.

With this established, the basic recursion formula for $\sigma(G)$ is readily derived. Consider the relation of the stable sets of G to those of G with a vertex p removed. Every stable set of $G - p$ is also a stable set of G. The other stable sets of G arise from unions of p with stable sets of $G - p$ which contain no vertex adjacent to p; the number of such sets is $\sigma(G - p - A_p)$, where A_p is the set of vertices of G adjacent to p. Thus the recursion formula for $\sigma(G)$ is

$$\sigma(G) = \sigma(G - p) + \sigma(G - p - A_p). \tag{1}$$

Suppose G is a two-component graph $G_1 \cup G_2$. Then the stable sets of G arise from unions of any stable set of G_1 with any of G_2. The second basic formula follows by induction:

$$\sigma\left(\bigcup_i G_i\right) = \prod_i \sigma(G_i). \tag{2}$$

One final result will complete the machinery for calculating $\sigma(G)$ for any molecular graph. Consider a linear chain of n atoms, the *path* P_n, and the effect of adding a point to one end to produce P_{n+1}. The recursion formula (1) gives

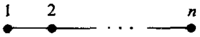
$$\sigma(P_{n+1}) = \sigma(P_n) + \sigma(P_{n-1}), \tag{3}$$

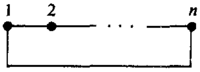
which is identical to the recursion relation for the Fibonacci numbers

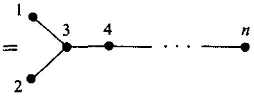
$$F_{n+1} = F_n + F_{n-1}, \qquad F_0 = F_1 = 1. \tag{4}$$

Table 1. Fibonacci numbers and $\sigma(G)$ for paths, cycles and *iso*-paths

n	F_n	$\sigma(P_n)^a$	$\sigma(C_n)^b$	$\sigma(i-P_n)^c$
0	1	1		
1	1	2		
2	2	3	3	
3	3	5	4	
4	5	8	7	9
5	8	13	11	14
6	13	21	18	23
7	21	34	29	37
8	34	55	47	60
9	55	89	76	97
10	89	144	123	157
11	144	233	199	254
12	233	377	322	411
13	377	610	521	665
14	610	987	843	1,076
15	987	1,597	1,364	1,741
16	1,597	2,584	2,207	2,817
17	2,584	4,181	3,571	4,558
18	4,181	6,765	5,778	7,375
19	6,765	10,946	9,349	11,933
20	10,946	17,711	15,127	19,308

[a] $P_n =$ 

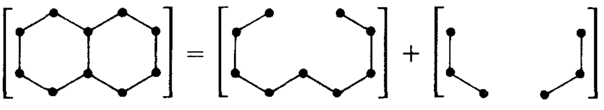
[b] $C_n =$ 

[c] $i - P_n =$ 

Since P_0 has one stable set ($\varnothing$) and P_1 has two ($\varnothing$, {1}), i.e., $\sigma(P_0) = 1$ and $\sigma(P_1) = 2$, we have

$$\sigma(P_n) = F_{n+1}. \tag{5}$$

The general strategy for evaluating $\sigma(G)$ is to use the recursion formula (1) as often as needed to eliminate all trivalent vertices from G and then to evaluate the resulting collection of paths and unions of paths with the help of (2) and (5). (Table 1 lists the first 20 Fibonacci numbers along with $\sigma(G)$ for three simple homologous series of molecules.) This procedure can be illustrated with the example of naphthalene, where $\sigma(G)$ is denoted by enclosing the graph in square brackets.



$$\begin{aligned} &= \sigma(P_9) + \sigma(P_3 \cup P_3) \\ &= \sigma(P_9) + [\sigma(P_3)]^2 \\ &= F_{10} + F_4^2 \\ &= 114 \end{aligned}$$

In order to evaluate the contributions of individual atoms and bonds to the spaces of the graph topology, a procedure for enumerating the number of open or closed sets which contain a given subset $A \subset X$ is required. To accomplish this we make use of the fact (cf. Sect. 2) that the collection of open sets containing A is isomorphic to the subspace topology of $X - \hat{A}$ so that, in particular, they contain the same number of sets. Similarly the number of closed sets containing A is equal to the number of sets in the subspace topology of $X - \bar{A}$.

Considering first the number of open or closed sets containing a given atom, recall from Sect. 3 that for a given orientation D of G each atom is either open or closed. If the atom p is open, then

$$\widehat{\{p\}} = \{p\}, \overline{\{p\}} = \{p\} \cup A_p, \tag{6}$$

and the number of open and of closed sets containing p are

$$\begin{aligned} n_o(p) &= \sigma(G - p), \qquad \text{and} \\ n_c(p) &= \sigma(G - p - A_p), \end{aligned} \tag{7}$$

respectively. If q is closed, one finds similarly

$$\begin{aligned} n_o(q) &= \sigma(G - q - A_q), \\ n_c(q) &= \sigma(G - q). \end{aligned} \tag{8}$$

Note that in either case, from (1) one has

$$n_o(p) + n_c(p) = \sigma(G).$$

Enumeration of the number of open or closed sets containing a given bond (p, q) proceeds similarly. If p is the open member of the pair,

$$\begin{aligned} \widehat{\{p, q\}} &= \{p\} \cup \{q\} \cup A_q \\ &= \{q\} \cup A_q, \end{aligned} \tag{9}$$

where the second equality follows from the fact that $p \in A_q$. Similarly

$$\begin{aligned} \overline{\{p, q\}} &= \{p\} \cup A_p \cup \{q\} \\ &= \{p\} \cup A_p. \end{aligned} \tag{10}$$

Therefore

$$\begin{aligned} n_o(p, q) &= \sigma(G - q - A_q), \\ n_c(p, q) &= \sigma(G - p - A_p), \end{aligned}$$

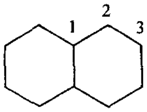
and the total number of open or closed sets containing (p, q) is

$$\begin{aligned} n_{pq} &= n_o(p, q) + n_c(p, q) \\ &= \sigma(G - p - A_p) + \sigma(G - q - A_q). \end{aligned} \tag{11}$$

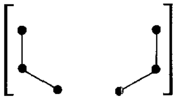
It is convenient to define $\alpha_p = \sigma(G - p - A_p)$ so that (11) becomes

$$n_{pq} = \alpha_p + \alpha_q. \tag{12}$$

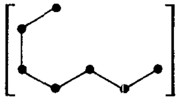
Continuing with the example of naphthalene



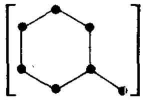
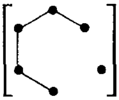
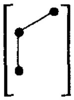
one finds

$\alpha_1 =$ 

$= F_4^2 = 25,$

$\alpha_2 =$ 

$= F_8 = 34,$

$\alpha_3 =$  $=$  $+$ 

$= F_6F_2 + F_4 = 31,$

and thus the four inequivalent bonds are ordered as follows:

$$n_{23} = 65,\ n_{3'3} = 62,\ n_{12} = 59,\ n_{1'1} = 50,$$

which is exactly the same as the ordering given by the bond orders p_{ij} from molecular orbital theory:

$$p_{23} = 0.725,\ p_{3'3} = 0.603,\ p_{12} = 0.555,\ p_{1'1} = 0.518.$$

5. Properties of the Function σ(G) and Related Quantities

The value of $\sigma(G)$ turns out to be quite sensitive to the details of molecular structure,[2] particularly to the extents of branching and cyclization. This is illustrated in Table 2, which lists $\sigma(G)$ for all six- and eight-atom π-graphs (i.e., graphs representing the carbon skeletons of conjugated π-electron systems) and in Table 3, which gives $\sigma(G)$ for the carbon skeletons of all alkanes from C_5 through C_8. It may be seen that this function does a good job of distinguishing among these structures although it does not quite reach the ideal of assigning a different number to every distinct structure. Inspection of the tables also shows that $\sigma(G)$ increases with branching and decreases with cyclization.

---

[2] The function $\sigma(G)$ is similar in this respect to Hosoya's (Ref. [3]) "topological index", which is equal to the number of stable *edge* sets of G, i.e., it is σ of the line graph of G.

Table 2. $\sigma(G)$ for alternant π-graphs

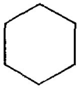
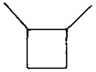
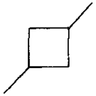
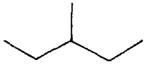
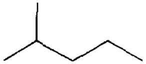
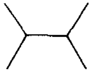
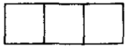
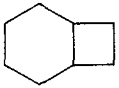
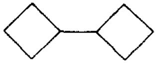
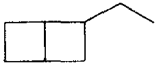
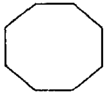
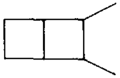
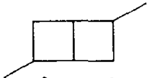
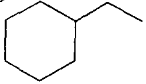
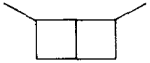
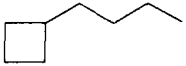
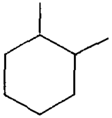
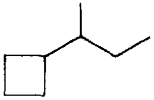
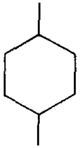
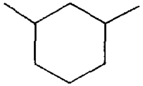
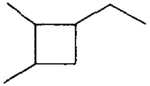
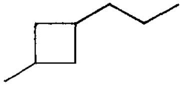
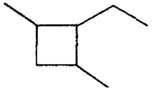
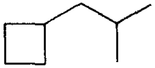
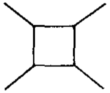
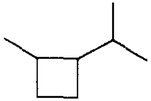
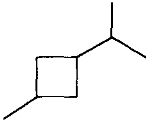
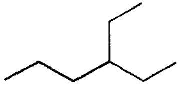
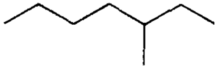
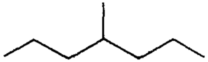
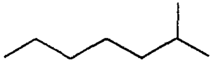
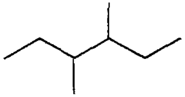
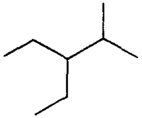
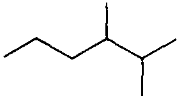
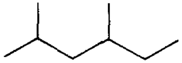
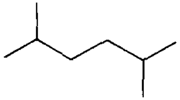
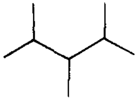
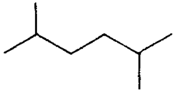
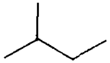
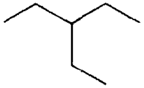
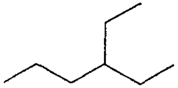
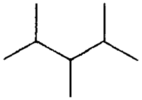
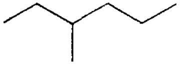
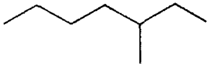
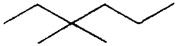
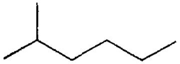
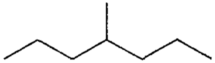
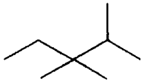
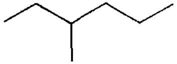
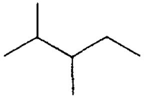
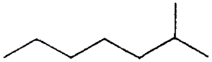
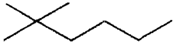
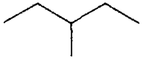
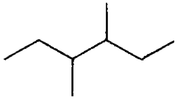
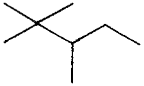
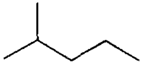
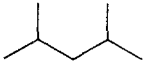
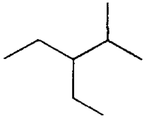
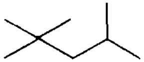
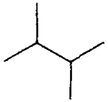
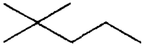
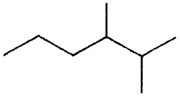
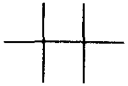
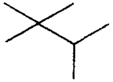
$n = 6$								
$\sigma(G)$:	17	18	19	20	21	22	23	24
G:					 			
$n = 8$								
$\sigma(G)$:	41	44	45	46	47	48	49	50
G:							 	 
$\sigma(G)$:	52	53	54	55	56	57	58	
G:	  		  	  	 	 		
$\sigma(G)$:	59	60	61	62	64	65	66	
G:		 						

Table 3. $\sigma(G)$ for C_5 to C_8 alkanes

	Molecule	σ		Molecule	σ		Molecule	σ	Molecule	σ
C_5:		13	C_7:		34	C_8:		55		65
		14			35			57		66
		17			36			58		66
					37			59		68
C_6:		21			38			60		69
		22			40			60		70
		23			41			61		77
		24			43			62		80
		26			44			62		

The latter observation can be made more quantitative as follows. Consider the process of adding to a graph G the edge e_{pq} joining vertices p and q which are not adjacent in G. From the basic recursion relation (1) we have

$$\sigma(G + e_{pq}) = \sigma(G - p) + \sigma(G - p - q - A_p), \tag{13}$$

and

$$\begin{aligned}\sigma(G) &= \sigma(G - p) + \sigma(G - p - A_p)\\ &= \sigma(G - p) + \sigma(G - p - q - A_p) + \sigma(G - p - q - A_p - A_q).\end{aligned} \tag{14}$$

Combining (13) and (14) yields

$$\sigma(G + e_{pq}) = \sigma(G) - \sigma(G - p - q - A_p - A_q), \tag{15}$$

which shows that adding a new edge to a graph always decreases $\sigma(G)$.

Remarkably, among the alkanes $\sigma(G)$ appears to be sensitive to the same aspects of molecular structure as are some of their physical properties. Fig. 2 shows the empirical correlation between $\sigma(G)$ and heats of formation and Fig. 3 illustrates a similar correlation with boiling points. In both cases the correlations are quite good, particularly for the lower alkanes. This is perhaps not too surprising, since it is known from empirical studies [1] that these physical properties are highly correlated with molecular branching.

In the two examples of butadiene and naphthalene given earlier it was found that the quantities n_{ij}, the total number of open and of closed sets containing the bond

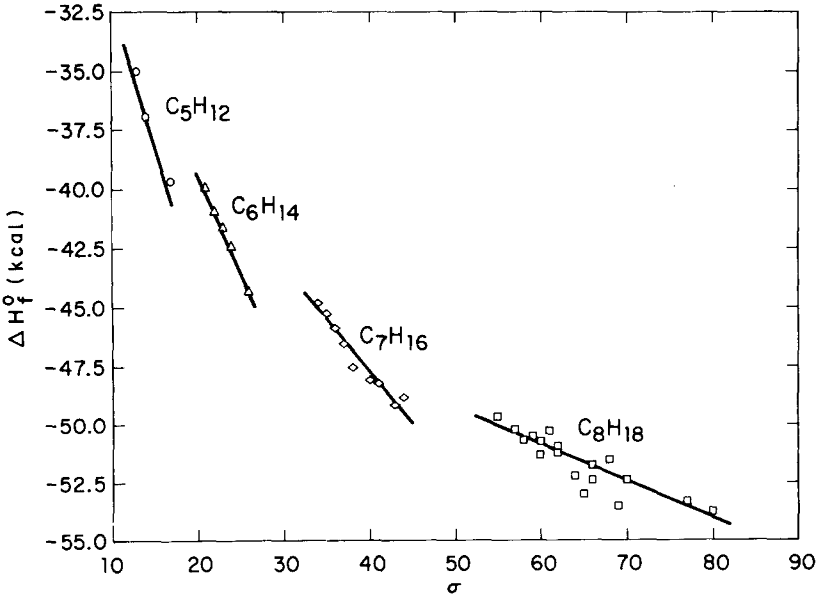


Fig. 2. Alkane heats of formation vs. $\sigma(G)$. The straight lines serve merely to group the points and have no theoretical significance

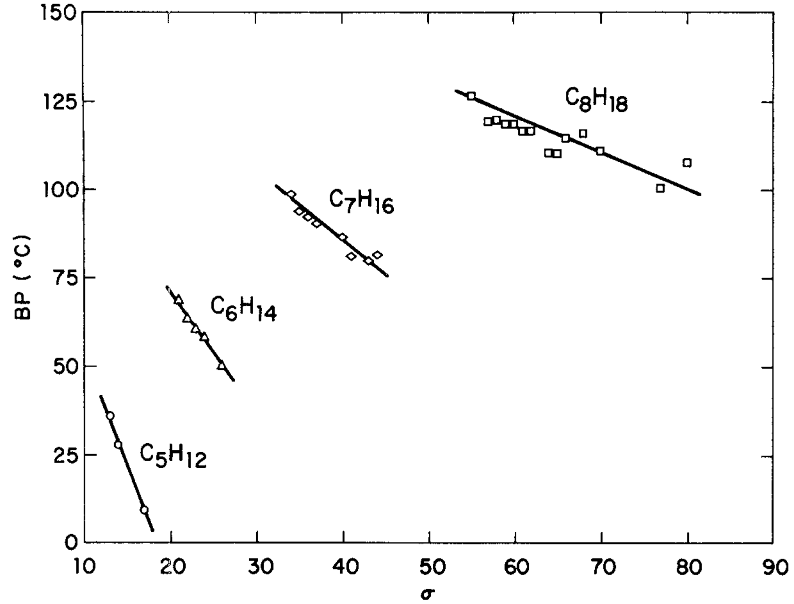


Fig. 3. Alkane boiling points vs. $\sigma(G)$. The straight lines serve merely to group the points and have no theoretical significance

(i, j), are an excellent measure of π-bond strengths in that they yield the same ordering as do the bond orders calculated from molecular orbital theory. Table 4 lists the values of n_{ij} and π-bond orders for a collection of 77 distinct bonds in 22 molecules. The intramolecular correlation between the topological and quantum-mechanical measures of bond strength is almost perfect. (Bonds ordered differently by these two measures are indicated by an asterisk in the table.) Even such features as bond alternation in linear polyenes and accidentally, i.e., non-symmetry-dictated, equal bond orders are accurately reflected by the topological measure.

In order to compare bonds in different molecules it is necessary to normalize the n_{ij} in some manner. The most obvious course is to employ the quantity n_{ij}/σ as a measure of bond strength. Fig. 4 shows, for the molecules in Table 4, the π-bond order plotted against n_{ij}/σ. The intermolecular correlation of these two quantities is fairly good but definitely inferior to the intramolecular correlation. We have examined a number of two-parameter forms as empirical expressions of this correlation; a convenient expression which works best for $p \geqslant 0.6$ is

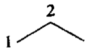
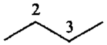
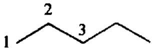
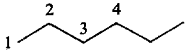
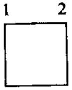
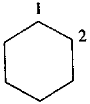
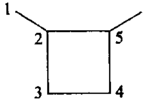
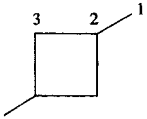
$$p_{ij} = (3n_{ij}/2\sigma)^{9/4},$$

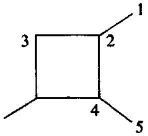
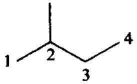
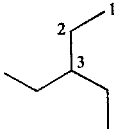
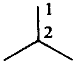
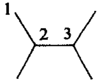
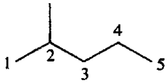
however the best overall fit is

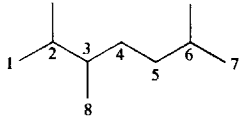
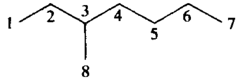
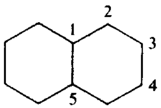
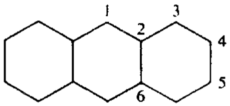
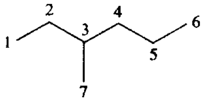
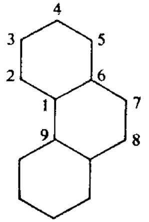
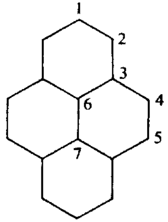
$$p_{ij} = \exp 3.81(n_{ij}/\sigma - 0.67), \tag{16}$$

for which the correlation coefficient is $r^2 = 0.86$. This function is plotted in Fig. 4.

Table 4. Topological parameters and bond orders for π-electron systems

Molecule	σ	(i, j)	n_{ij}	p_{ij}
	3	1, 2	2	1.000
	5	1, 2	3	0.707
	8	1, 2	5	0.894
		2, 3	4	0.447
	13	1, 2	8	0.789
		2, 3	7	0.577
	21	1, 2	13	0.871
		3, 4	12	0.785
		2, 3	11	0.483
	7	1, 2	4	0.707
	18	1, 2	10	0.667
	20	1, 2	12	0.871
		3, 4	12	0.785
		2, 3	10	0.388
		2, 5	8	0.301
	21	1, 2	12	0.724
		2, 3	11	0.447

Molecule	σ	(i, j)	n_{ij}	p_{ij}
	34	4, 5	21	0.881
		1, 2	20	0.756
		2, 3	18	0.513
		2, 4	14	0.325
	14	3, 4	9	0.924
		1, 2	8	0.653
		2, 3	6	0.383
	35	1, 2	22	0.833
		2, 3	17	0.500
	9	1, 2	5	0.577
	24	1, 2	14	0.667
		2, 3	8	0.333
	23	4, 5	14	0.761
		1, 2	13	0.616
		3, 4	13	0.616
		2, 3	11	0.470

	176	4, 5	106	0.826
		3, 8	106	0.826
		1, 2	102	0.661
		6, 7	100	0.645
		3, 4	86	0.439
		5, 6	80	0.406
		2, 3	64	0.349
	58	1, 2	37	0.911
		6, 7	36	0.874
		3, 8	34	0.798
		4, 5	34	0.798
		5, 6	30	0.479
		3, 4	28	0.441
		2, 3	26	0.404
	114	2, 3	65	0.725
		3, 4	62	0.603
		1, 2	59	0.555
		1, 5	50	0.518
	726	3, 4	415	0.737
		1, 2*	388	0.606
		4, 5	394	0.586
		2, 3	373	0.535
		2, 6	310	0.485
	36	1, 2	23	0.916
		5, 6	22	0.782
		3, 7	21	0.723
		4, 5	20	0.590
		3, 4	18	0.531
		2, 3	16	0.397
	717	7, 8	418	0.775
		4, 5	409	0.707
		2, 3	406	0.702
		3, 4	391	0.623
		1, 2	379	0.590
		5, 6	370	0.575
		1, 6*	325	0.542
		6, 7	364	0.506
		1, 9	340	0.461
	1733	4, 5	1016	0.777
		1, 2	978	0.670
		2, 3	894	0.594
		6, 7	890	0.536
		3, 6*	808	0.524
		3, 4	871	0.505

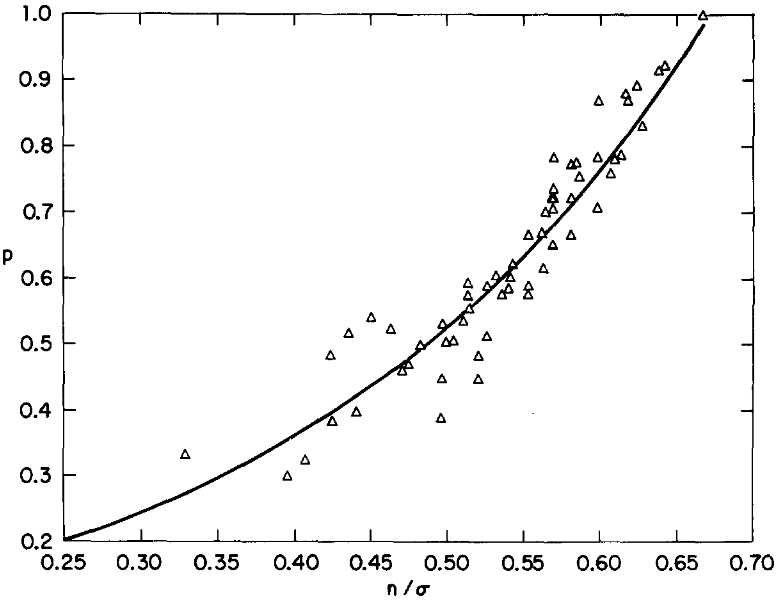


Fig. 4. Hückel bond order vs. n_{ij}/σ for the bonds in Table 4. The equation of the solid curve is $p = \exp 3.81\ (n/\sigma - 0.67)$

From the scatter of the points evident in the figure, it is unlikely that any other expression will do significantly better.[3]

6. Summary

Topology provides the necessary and appropriate tools for discussing the nonmetric aspects of molecular structure. It is as useful in probing the structures of the spaces of finitely many points which arise naturally in this context as it is in its more usual application to continuous spaces, although the details of the analysis are quite different in these two cases.

There is an essentially unique pair of topological spaces associated with any alternant molecular framework.[4] This association can be regarded either as a consequence of the fact that any finite topological space induces a unique graph whose vertex set consists of the points of the space or as a result of partitioning the real space occupied by a molecule into atomic subregions which become the points of the resulting quotient space. Many concepts from general topology retain their useful-

---

[3] Note the contrast to the situation in the bond topology for which bond orders are *linearly* correlated with the sum of the *cardinalities* of the closed sets containing the bond. The very different measures required in these two cases reflect the profoundly different structures of the topological spaces.

[4] That this is not so for nonalternants points up the essentially topological distinction between these two classes of structures. As yet we have been unable to discover a "natural" way of topologizing nonalternants.

ness and intuitive content in these molecular topological spaces. For example, a connected topological space corresponds to a connected molecule and the only molecules which are homeomorphic are stereoisomers. However, finite topological spaces differ from continuous spaces in that many of the most important questions regarding their structures are combinatorial in nature, that is, questions as to the existence and number of subsets satisfying certain conditions. Thus, for example, the number of open sets in a molecular space is a measure of molecular complexity and the number of open sets and closed sets containing a pair of adjacent atoms is a measure of the strength of the bond between them. These combinatorial questions turn out to be highly amenable to solution by graph-theoretical methods – another consequence of the intimate association between finite topological spaces and graphs – and the necessary mathematical machinery to this end has been devised.

Some of the quantities resulting from analysis of the structure of these spaces turn out to be well correlated with physical properties such as heats of formation and bond strengths. From this one might hypothesize that such properties are in large part determined by the molecular topology, although at present we are unable to show precisely how or where the topology constrains the physics.

References

1. Kier, L. B., Hall, L. H.: Molecular connectivity in chemistry and drug research. New York: Academic Press 1976
2. Balaban, A. T., (Ed.): Chemical applications of graph theory. New York: Academic Press 1976
3. Hosoya, H., Hosoi, K.: J. Chem. Phys. **64**, 1065 (1976); Hosoya, H., Hosoi, K., Gutman, I.: Theoret. Chim. Acta (Berl.) **38**, 37 (1975); Hosoya, H., Murakami, M.: Bull. Chem. Soc. Japan **48**, 3512 (1975); Hosoya, H.: Bull. Chem. Soc. Japan **44**, 2332 (1971)
4. Simmons, H. E., Merrifield, R. E.: Proc. Natl. Acad. Sci. **74**, 2616 (1977)
5. Two texts which we have found particularly useful are: Munkries, J. R.: Topology: A first course. Englewood Cliffs, N.J.: Prentice-Hall 1975; Dugundji, J.: Topology. Boston: Allyn and Bacon 1966
6. Sharp, H., Jr.: Proc. Am. Math. Soc. **17**, 1344 (1966)
7. Evans, J. W., Harary, F., Lynn, M. S.: Commun. ACM **10**, 295 (1967)

Received June 18, 1979