

TWO METRICS FOR A GRAPH-THEORETICAL MODEL OF ORGANIC CHEMISTRY

V. KVASNIČKA and J. POSPÍCHAL

Department of Mathematics, Faculty of Chemistry, Slovak Technical University, 81237 Bratislava, Czechoslovakia

Received 9 May 1988

(in final form 23 January 1989)

**Dedicated to the memory of
Professor Milan Sekanina**

Abstract

A graph-theoretical model of organic chemistry is proposed. The main idea behind this model is a molecular graph in the form of a multigraph with loops; its vertices are evaluated by vertex labels (atomic symbols). The chemical distance between two graphs from the same family of isomeric graphs is based on the maximal common subgraph. The produced reaction graph is composed of the minimal number of edges and/or loops. The reaction distance assigned to the chemical transformation $G_1 \rightarrow G_2$ is equal to the minimal number of the so-called elementary transformations that are necessary for the transformation of G_1 into G_2 . Because these metrics are not "isometric", the resulting reaction graphs may depend on the metric used.

1. Introduction

The purpose of the present communication is to suggest and elaborate a mathematical model of organic chemistry following the concepts and notions of graph theory [1–6]. The model is focused toward the theory of organic synthesis design [7–14]. It provides simple, unambiguous and effective mathematical tools for the formalization of many phenomena of organic chemistry, e.g. structural formulas, chemical reactions, mechanisms of reactions, and so on. In our previous communications [15–18], we have devised a mathematical model of organic chemistry based on a graph-theoretic formalism; the theory developed was fully compatible with the matrix model of Dugundji and Ugi [8,9]. The present approach extends our initial ideas on the application of graph theory for the purposes of the formalization of

organic synthesis design. In particular, two different metrics for the families of isomeric molecular graphs are suggested. The chemical metric [19,20] based on the concept of the maximal common subgraph provides reaction graphs [17] that are consistent with Ugi et al.'s principle of minimal chemical distance [9,21].

The same metric was also simultaneously suggested by Johnson [22,23]. He introduced this very important concept for the study of structure-activity relationships. The concept of the maximal common subgraph has been advantageously used by Lynch and Willett [24–27] in a process of looking for the reaction centre assigned to a chemical reaction which transforms an educt molecule onto a product molecule. Recently, Ugi et al. [28–30] have published a series of papers that are also devoted to the graph-theoretical aspects of their principle of minimal chemical distance.

The second metric, called the reaction metric and initially introduced by Koča [31,32], enables a formalization of the chemical transformation $G_1 \rightarrow G_2$ (G_1 and G_2 are molecular graphs) into sequences of elementary transformations which reflect the mechanism of the given chemical transformation. Both metrics provide effective rules for constructing the reaction graphs and their decomposition into mechanisms. In order to keep the theory as simple as possible, we postulate that the molecules contain an even number of electrons and that all bonds are realized by two electrons (to some extent this is identical with the restricted chemistry of Dugundji and Ugi [8]).

2. Basic concepts

A *vertex set* $V = \{v_1, v_2, \dots, v_N\}$ is a nonempty set composed of N vertices (atoms) v_1, v_2, \dots, v_N . A *vocabulary* $\mathcal{V} = \{\alpha_1, \alpha_2, \dots, \alpha_Q\}$ is a nonempty set composed of Q *vertex labels* (atomic symbols) $\alpha_1, \alpha_2, \dots, \alpha_Q$. The vertex set V is *mapped* into the vocabulary \mathcal{V} ,

$$\varphi: V \rightarrow \mathcal{V}. \quad (2.1)$$

This means that each vertex is uniquely evaluated by a vertex label. An *edge* is an unordered pair of distinct vertices from the vertex set V . The edge $[v_i, v_j]$ is *incident* with the vertices v_i and v_j and connects them. Two distinct edges are *adjacent* if they have a vertex in common. Two distinct vertices are *adjacent* if they are incident with the same edge. A *multiedge* of multiplicity t is a set composed of t edges incident with the same pair of distinct vertices. We shall assume that the multiplicity of a multiedge is restricted to $1 \leq t \leq 3$, i.e. at most triple edges can appear in our graph-theoretical model. An *edge set* $E = \{e_1, e_2, \dots, e_M\}$, where $e_i = [v_j, v_k]$, associated with the vertex set V is a set of M edges e_1, e_2, \dots, e_M with the vertices in V .

A *loop* is the pair obtained by taking the same vertex twice from the vertex set V . The loop $[v_i, v_i]$ is *incident* with the vertex $v_i \in V$. A *multiloop* of multi-

plicity u is a set of u loops incident with the same vertex. We shall assume that the multiplicity u is restricted to $1 \leq u \leq 4$. A *loop set* $L = \{l_1, l_2, \dots, l_p\}$, where $l_i = [v_j, v_j]$, associated with the vertex set V is a set of P loops with the vertices from V .

Geometrically, the vertex $v_i \in V$ is represented by a heavy dot, the edge $[v_i, v_j] \in E$ is represented by a continuous line connecting the vertices v_i and v_j , and the loop $[v_i, v_i] \in L$ is conventionally represented by a continuous line beginning and ending at the same vertex v_i .

$$\dot{v} \Leftrightarrow \text{vertex } v \in V \quad (2.2a)$$

$$\overset{\bullet}{\underset{\bullet}{\text{---}}} \Leftrightarrow \text{edge } [v, v'] \in E \quad (2.2b)$$

$$\begin{array}{c} \text{O} \\ | \\ \bullet \end{array} \Leftrightarrow \text{loop } [v, v] \in L \quad (2.2c)$$

DEFINITION 2.1

A *graph* is an ordered triplet

$$G = (V, E, L), \quad (2.3)$$

where V is a vertex set. E and L are edge and loop sets, respectively, both associated with the vertex set V . Geometrically, the graph G is represented by heavy dots (vertices) and lines connecting two distinct vertices (edges) or beginning and ending at the same vertex (loops).

DEFINITION 2.2

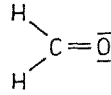
A *molecular graph* is an ordered 5-tuple

$$G_M = (V, E, L, \varphi, \mathcal{V}), \quad (2.4)$$

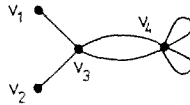
where V is a vertex set, E and L are edge and loop sets, respectively, both associated with the vertex set V , and φ is a mapping (2.1) of the vertex set into the vocabulary \mathcal{V} of vertex labels. We say that the molecular graph $G_M = (V, E, L, \varphi, \mathcal{V})$ is *induced* by the graph $G = (V, E, L)$.

EXAMPLE 2.1

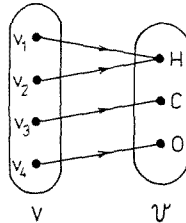
The formaldehyde molecule



is represented by the graph $G = (V, E, L)$



where $V = \{v_1, v_2, v_3, v_4\}$, $E = \{e_1 = [v_1, v_3], e_2 = [v_2, v_3], e_3 = [v_3, v_4], e_4 = [v_3, v_4]\}$. $L = \{l_1 = [v_4, v_4], l_2 = [v_4, v_4]\}$. The vertex set V is composed of four vertices, the edge set E is composed of four edges, where the edges e_3 and e_4 represent a multiedge of double multiplicity, and the loop set L is composed of two loops l_1 and l_2 which form a multiloop with double multiplicity. This graph induces the molecular graph $G_M = (V, E, L, \varphi, \mathcal{V})$ and the mapping φ into the vocabulary \mathcal{V} is determined as follows:



From the graph $G = (V, E, L)$ we may form its *subgraph*, which is also a graph, by successive applications of one of the following two processes:

- (1) Deleting a preselected vertex and all edges/loops that are incident with the vertex.
- (2) Deleting a preselected edge/loop, where the vertices incident with the edge/loop remain in the graph.

DEFINITION 2.3

A graph $G' = (V', E', L')$ is a *subgraph* of the graph $G = (V, E, L)$ iff $V' \subseteq V$, $E' \subseteq E$, and $L' \subseteq L$. If the subgraph was formed from the graph G by making use of merely the first process (the vertices were deleted), then the subgraph G' is called the *subgraph induced* by the vertex set V' . The property of being the subgraph is denoted by $G' \subseteq G$.

DEFINITION 2.4

A molecular graph $G'_M = (V', E', L', \varphi', \mathcal{V})$ is a subgraph of the molecular graph $G_M = (V, E, L, \varphi, \mathcal{V})$ iff $G' = (V', E', L') \subseteq G = (V, E, L)$, where the graph $G'(G)$ induces the molecular graph $G'_M(G_M)$, and the mapping $\varphi': V' \rightarrow \mathcal{V}$ is a restriction of $\varphi: V \rightarrow \mathcal{V}$.

A union of two molecular graphs $G'_M = (V', E', L', \varphi', \mathcal{V})$ and $G''_M = (V'', E'', L'', \varphi'', \mathcal{V})$ determined over disjoint vertex sets is defined by

$$G_M = G'_M \cup G''_M = (V = V' \cup V'', E = E' \cup E'', L = L' \cup L'', \varphi, \mathcal{V}), \quad (2.5)$$

where the mapping φ is

$$\varphi(v) = \begin{cases} \varphi'(v) & \text{for } v \in V' \\ \varphi''(v) & \text{for } v \in V'' \end{cases} \quad (2.6a)$$

$$(2.6b)$$

The resulting molecular graph is termed *disconnected*, and its *components* are G'_M and G''_M , which are also the subgraphs of G_M . In the same way, a union of two graphs G' and G'' can also be determined; the mappings $\varphi', \varphi'', \varphi$ are now ignored.

One of the most fundamental notions of organic chemistry is the *isomerism*. Two or more molecules are *isomeric* iff they contain the same number of atoms of the same elements and the same number of valence electrons, but differ in the structural arrangement of the atoms. For graphs and molecular graphs, this notion can be simply represented.

DEFINITION 2.5

Two graphs $G_1 = (V_1, E_1, L_1)$ and $G_2 = (V_2, E_2, L_2)$ are isomeric, ($G_1 \approx G_2$), iff

$$|V_1| = |V_2| \quad (2.7a)$$

$$|E_1| + |L_1| = |E_2| + |L_2|, \quad (2.7b)$$

where $|X|$ denotes the number of elements (cardinality) of the set X .

In order to specify this definition of the isomerism for molecular graphs, we have to ensure the above condition that they contain the same number of vertices (atoms) of the same element [15].

DEFINITION 2.6

Two vertex sets $V_1 = \{v_1, v_2, \dots, v_N\}$ and $V_2 = \{v'_1, v'_2, \dots, v'_N\}$ with mappings $\varphi'_1 = V_1 \rightarrow \mathcal{V}$ and $\varphi_2 = V_2 \rightarrow \mathcal{V}$, respectively, are *similar* iff there exists a

one-to-one correspondence ψ between them, $\psi: V_1 \rightarrow V_2$, which conserves the evaluation of vertices, i.e. $\psi(v_i) = v'_i$ implies $\varphi_1(v_i) = \varphi_2(v'_i)$.

DEFINITION 2.7

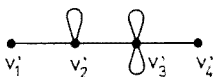
Two molecular graphs $G_{M,1} = (V_1, E_1, L_1, \varphi_1, \mathcal{V})$ and $G_{M,2} = (V_2, E_2, L_2, \varphi_2, \mathcal{V})$ are *isomeric*, ($G_{M,1} \approx G_{M,2}$), iff the graphs $G_1 = (V_1, E_1, L_1)$ and $G_2 = (V_2, E_2, L_2)$ inducing these molecular graphs are isomeric and the vertex sets V_1 and V_2 are similar.

EXAMPLE 2.2

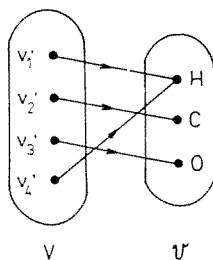
Let us consider the hydroxycarbene molecule



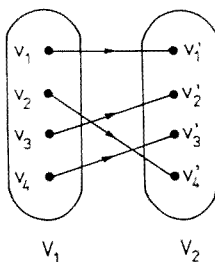
This molecule is isomeric with the formaldehyde molecule presented in example 2.1. The hydroxycarbene molecule is represented by the graph



where the mapping φ_2 is



The graph from example 2.1 is isomeric with the present graph. The mapping ψ is determined as follows



We see that mapping ψ conserves the evaluation of vertices by atomic symbols.

Let us study a molecular graph $G_M = (V, E, L, \varphi, \mathcal{V})$. We assign to this molecular graph the so-called *atomic set* $A = \{a_1, a_2, \dots, a_N\}$ composed of lexicographically ordered atomic symbols (ordered, for example, with respect to atomic numbers) $a_1 \leq a_2 \leq \dots \leq a_N$ of vertices from V , formally as $as(G_M) = A$. If two molecular graphs G_M and G'_M are isomeric, $G_M \approx G'_M$, then they have the same atomic sets as $as(G_M) = as(G'_M)$. For instance, the isomeric molecular graphs considered in examples 2.1 and 2.2 have the same atomic sets $A = \{H, H, C, O\}$.

DEFINITION 2.8

Two graphs $G_1 = (V_1, E_1, L_1)$ and $G_2 = (V_2, E_2, L_2)$ are *isomorphic*, ($G_1 \cong G_2$), iff there exists such one-to-one mapping $\psi: V_1 \rightarrow V_2$ which induces two one-to-one mappings $\chi: E_1 \rightarrow E_2$ and $\chi': L_1 \rightarrow L_2$ that conserve the incidences of edges and loops, respectively.

DEFINITION 2.9

Two molecular graphs $G_{M,1} = (V_1, E_1, L_1, \varphi_1, \mathcal{V})$ and $G_{M,2} = (V_2, E_2, L_2, \varphi_2, \mathcal{V})$ are *isomorphic*, ($G_{M,1} \cong G_{M,2}$), iff the graphs $G_1 = (V_1, E_1, L_1)$ and $G_2 = (V_2, E_2, L_2)$ inducing these molecular graphs are isomorphic and the mapping ψ (from definition 2.8) induces the similarity of the vertex sets V_1 and V_2 .

How can one interpret simply the notion of isomerism for molecular graphs? Loosely speaking, and with some simplification, the isomorphic molecular graphs $G_{M,1}$ and $G_{M,2}$ differ only in the indexing of their vertices (atoms). Since the physical and chemical properties of a molecule are invariant with respect to the indexing of its atoms, isomorphic molecular graphs are physically and chemically equivalent. A molecule can be represented by many different molecular graphs, but they must be strictly isomorphic. This means that the structural formulas (or molecular graphs) represent a molecule up to isomorphism.

The concept of isomerism determined over a universe of graphs can be formally considered as a realization of an "equivalence" relation between abstract elements. For instance, one can verify simply that the following three laws are satisfied:

- (1) Reflexive law, $G \cong G$.
- (2) Symmetric law, if $G_1 \cong G_2$, then $G_2 \cong G_1$.
- (3) Transitive law, if $G_1 \cong G_2$ and $G_2 \cong G_3$, then $G_1 \cong G_3$.

The same three laws are also satisfied for the notion of isomorphism. The universe of graphs can be decomposed onto disjoint families of mutually isomeric graphs.

DEFINITION 2.10

The family of isomeric graphs \mathcal{F}_{pq} is composed of all possible nonisomorphic graphs with p vertices and q edges and loops,

$$\mathcal{F}_{pq} = \{G = (V, E, L); |V| = p \text{ and } |E| + |L| = q\}. \quad (2.8)$$

The above definition can be simply extended for molecular graphs.

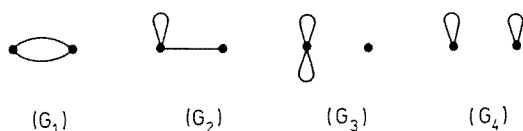
DEFINITION 2.11

The family $\mathcal{F}_{pq}(A)$ of molecular graphs is composed of all possible non-isomorphic molecular graphs with the same atomic set A and with p vertices and q edges and loops.

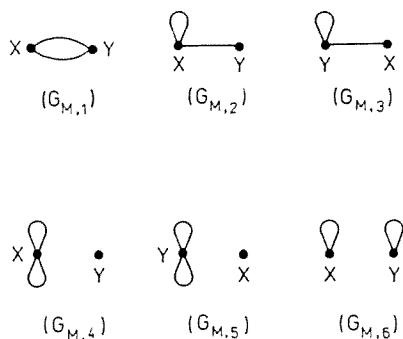
$$\mathcal{F}_{pq}(A) = \{G = (V, E, L); |V| = p \text{ and } |E| + |L| = q, \text{ and } \text{as}(G_M) = A\}. \quad (2.9)$$

EXAMPLE 2.2

The family $\mathcal{F}_{2,2}$ of isomeric graphs composed of two vertices and two edges and/or loops contains the following four graphs:



This family can be specified for the atomic set $A = X, Y$ in the form of the family $\mathcal{F}_{2,2}(A)$ of isomeric molecular graphs.



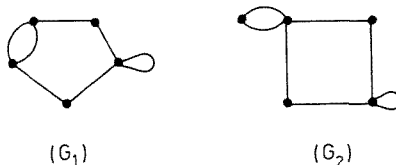
3. Chemical distance

For the mathematical modelling of organic chemistry, the notion of *chemical distance* was initially introduced by Dugundji and Ugi [8,9]. They defined the chemical distance between two molecules (molecular graphs) as the Hamming norm of the difference in their adjacency matrices. In our approach [14–18], based on graph theory, this representation of chemical distance cannot be used; an unambiguous determination of the adjacency matrix requires an indexing of the vertices.

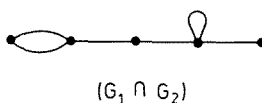
The notion of chemical distance will be determined for graphs. Its specification for molecular graphs is obvious and straightforward. Let us consider two isomeric graphs G_1 and G_2 , a *common subgraph* of these graphs is a graph G which is simultaneously isomorphic to the subgraphs $G'_1 \subseteq G_1$ and $G'_2 \subseteq G_2$, that is, $G \approx G'_1 \approx G'_2$. A *maximal common subgraph* of G_1 and G_2 , denoted by $G_1 \cap G_2$, is the common subgraph which contains the largest possible number of edges and loops.

EXAMPLE 3.1

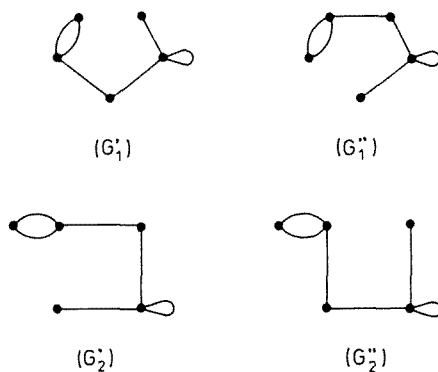
Consider the following two graphs:



The maximal common subgraph is

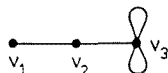


This graph is isomorphic to two subgraphs of G_1 and two subgraphs of G_2 :

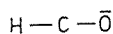


EXAMPLE 3.2

The maximal common subgraph of molecular graphs from examples 2.1 and 2.2 is



with mapping $\varphi(v_1) = \text{H}$, $\varphi(v_2) = \text{C}$, and $\varphi(v_3) = \text{O}$. This means, loosely speaking, that the maximal common substructure (fragment) of formaldehyde and hydroxycarbene molecules is



DEFINITION 3.1

The *chemical distance* [19,20] between two isomeric graphs $G_1 = (V_1, E_1, L_1)$ and $G_2 = (V_2, E_2, L_2)$ with maximal common subgraph $\tilde{G} = G_1 \cap G_2 = (V_{12}, E_{12}, L_{12})$ is given by

$$d(G_1, G_2) = |E_1| + |E_2| + |L_1| + |L_2| - 2|E_{12}| - 2|L_{12}|. \quad (3.1)$$

The chemical distance between a pair of isomeric graphs corresponds to the number of edges and loops that cannot be matched in the construction of a maximal common subgraph. It enables one to "measure" the similarity (or dissimilarity) of two different isomeric graphs; its increasing value indicates that the graphs are more dissimilar. We emphasize that the chemical distance for two isomeric molecular graphs is determined in a quite similar way to definition 3.1 for graphs.

EXAMPLE 3.3

Evaluate the chemical distance for G_1 and G_2 from example 3.1. These graphs are isomeric, i.e. $|E_1| + |L_1| = |E_2| + |L_2| = 6 + 1 = 7$. The maximal common subgraph satisfies $|E_{12}| = 5$ and $|L_{12}| = 1$. Hence, from example 3.1, the chemical distance is $d = 2$.

The chemical distance between molecular graphs of formaldehyde and hydroxycarbene for their maximal common subgraph specified in example 3.2 is $d = 4$.

THEOREM 3.1

The chemical distance $d(G_1, G_2)$ for isomeric graphs from the family \mathcal{F}_{pq} is a *metric* [19,20], and the following three properties are satisfied:

- (1) Positive semidefiniteness,

$$d(G_1, G_2) \geq 0 \quad (= 0 \text{ only for } G_1 \approx G_2). \quad (3.2a)$$

(2) Symmetry,

$$d(G_1, G_2) = d(G_2, G_1). \quad (3.2b)$$

(3) Triangle inequality,

$$d(G_1, G_2) + d(G_2, G_3) \geq d(G_1, G_3). \quad (3.2c)$$

The first two properties of the chemical distance result immediately from definition 3.1. The triangle inequality (3.2c) was proved using two different approaches [19,20]; here we present a simplified version of one of them [20].

Let us define, for a pair of isomeric graphs $G_1 = (V_1, E_1, L_1)$ and $G_2 = (V_2, E_2, L_2)$, the following auxiliary function:

$$g(G_1, G_2) = |E_1| - |E_{12}| + |L_1| - |L_{12}|. \quad (3.3)$$

This expresses the number of edges and/or loops that must be deleted from G_1 to obtain a subgraph which is isomorphic to the maximal common subgraph $G_1 \cap G_2$. The distance (3.1) is rewritten in a manifestly symmetric form (cf. (3.2b)):

$$d(G_1, G_2) = g(G_1, G_2) + g(G_2, G_1). \quad (3.4)$$

The following simple properties of the function g are easily proved:

$$(1) \quad G_1 \supseteq G_2 \Rightarrow g(G_1, G_2) = |E_1| + |L_1| - |E_2| - |L_2| \quad (3.5a)$$

$$(2) \quad g(G_1, G_2) = g(G_1, G_1 \cap G_2) \quad (3.5b)$$

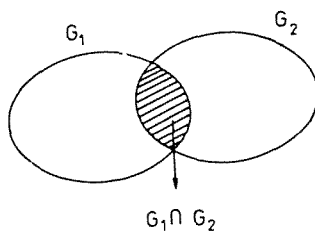
$$(3) \quad G'_1 \subseteq G_1 \Rightarrow g(G'_1, G_2) \leq g(G_1, G_2) \quad (3.5c)$$

$$(4) \quad G'_2 \supseteq G_2 \Rightarrow g(G_1, G'_2) \leq g(G_1, G_2) \quad (3.5d)$$

$$(5) \quad G_1 \subseteq G_2 \subseteq G_3 \Rightarrow g(G_1, G_2) + g(G_2, G_3) = g(G_1, G_3) \quad (3.5e)$$

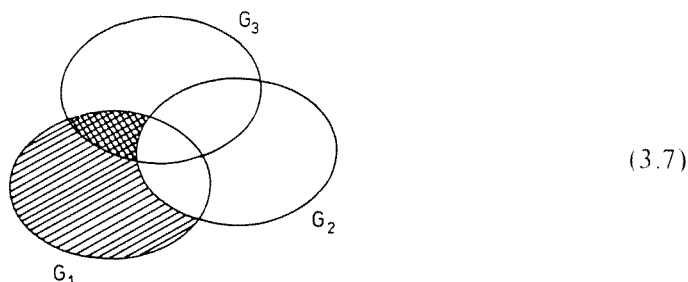
$$(6) \quad g(G_1, G_2) \geq g(G_1 \cap G_3, G_2). \quad (3.5f)$$

All these relations have a very simple "heuristic" set-theoretical interpretation: the graphs G_1 and G_2 are interpreted as sets; their maximal common subgraph may then be formally expressed as an intersection of "sets" G_1 and G_2 ,



(3.6)

The function $g(G_1, G_2)$ is equal to the cardinality of "subset" $G_1 - G_1 \cap G_2$ (unshaded area in G_1). This simple set-theoretical approach will be applied to prove property (3.5f): its Venn diagram has the form



The function $g(G_1, G_2)$ is equal to the cardinality of the single- and double-shaded area, whereas the function $g(G_1 \cap G_3, G_2)$ is equal to the cardinality of the double-shaded area, and thus the inequality (3.5f) is satisfied.

In order to prove the triangle inequality (3.2c), it is fully sufficient to verify the following two independent inequalities:

$$g(G_1, G_2) + g(G_2, G_3) \geq g(G_1, G_3) \quad (3.8a)$$

$$g(G_2, G_1) + g(G_3, G_2) \geq g(G_3, G_1). \quad (3.8b)$$

Using properties (3.5a–f), the left-hand side of (3.8a) is successively rewritten as follows:

$$\begin{aligned} g(G_1, G_2) + g(G_2, G_3) &\geq g(G_1, G_1 \cap G_2) + g(G_1 \cap G_2, G_3) \\ &= g(G_1, G_1 \cap G_2) + g(G_1 \cap G_2, G_1 \cap G_2 \cap G_3) \\ &= g(G_1, G_1 \cap G_2 \cap G_3) \geq g(G_1, G_1 \cap G_3) \\ &= g(G_1, G_3), \end{aligned} \quad (3.9)$$

where $G_1 \cap G_2 \cap G_3$ is a maximal common subgraph of G_1, G_2 and G_3 . The second inequality (3.8b) is verified in a similar way. Thus, we have proved the triangle inequality (3.2c).

We should stress that the chemical distance was defined via the notion of the maximal common subgraph. Hence, in order to evaluate the chemical distance $d(G_1, G_2)$ we have to know their maximal common subgraph $G_1 \cap G_2$. McGregor [33] has suggested a back-track searching algorithm for the construction of the maximal common subgraph of two graphs. His method involves, at worst, $N!$ back-track searches.

For molecular graphs, where the vertices are evaluated by chemical symbols, this approach could be substantially accelerated by making use of some effective heuristics.

Ugi et al. [21] have solved the same problem reformulated in the algebraic form. Let A_1 and A_2 be the adjacency matrices of the graphs G_1 and G_2 , respectively: the chemical distance between these graphs may then be alternatively defined as:

$$d(G_1, G_2) = \min \|A_1 - P^T A_2 P\|, \quad (3.10)$$

where $\|A\| = \sum_{i \leq j} |a_{ij}|$ is the Hamming norm of the symmetric matrix $A = (a_{ij})$ and P is the so-called permutation matrix. Ugi et al. [21] showed that the minimization of (3.10) corresponds to an integer programming problem.

Assuming that the graphs are taken from the same family \mathcal{F}_{pq} of isomeric graphs, then the following theorem is satisfied:

THEOREM 3.2

The chemical distance $d(G_1, G_2)$ for $G_1, G_2 \in \mathcal{F}_{pq}$ is even and bounded from above,

$$d(G_1, G_2) = 2k \leq 2q, \quad (3.11)$$

where $k = 0, 1, 2, \dots, q$.

Since $G_1 = (V_1, E_1, L_1), G_2 = (V_2, E_2, L_2)$ are taken from the same family \mathcal{F}_{pq} ,

$$|E_1| + |L_1| = |E_2| + |L_2| = q.$$

Introducing this relation into (3.1), we obtain

$$d(G_1, G_2) = 2(q - |E_{12}| - |L_{12}|),$$

as was to be proved.

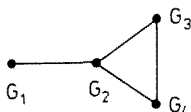
For a fixed family \mathcal{F}_{pq} of isomeric graphs, we construct the so-called *graph of chemical distances*, denoted by $\mathcal{G}_{pq}^{\text{CD}}$. Its vertex set is identified with the family \mathcal{F}_{pq} , i.e. each vertex of $\mathcal{G}_{pq}^{\text{CD}}$ corresponds formally to some graph of the family \mathcal{F}_{pq} . Two distinct vertices are connected by an edge iff the chemical distance between the corresponding graphs is equal to 2.

EXAMPLE 3.4

The matrix of chemical distances between graphs from the family $\mathcal{F}_{2,2}$ (see example 2.3) is a symmetric matrix with diagonal entries equal to zero,

$$d = \begin{pmatrix} 0 & 2 & 4 & 4 \\ 2 & 0 & 2 & 2 \\ 4 & 2 & 0 & 2 \\ 4 & 2 & 2 & 0 \end{pmatrix}$$

The graph of chemical distances constructed over the family $\mathcal{F}_{2,2}$ contains four vertices and four edges.



We see, for instance, that the chemical distance between graphs G_1 and G_3 is equal to twice their graph distance, where by the term "graph distance" [1] we understand the length (number of edges) of a minimal path between the vertices.

THEOREM 3.3

The chemical distance between graphs from the family \mathcal{F}_{pq} is equal to twice the graph distance between them in the graph \mathcal{G}_{pq}^{CD} .

Let us consider, for simplicity, a pair of graphs $G_1, G_2 \in \mathcal{F}_{pq}$ with chemical distance $d(G_1, G_2) = 4$. This means that two edge/loops in G_1 as well as in G_2 should be removed to get a subgraph isomorphic to their maximal common subgraph $G_1 \cap G_2$. From the graphs G_1 and G_2 we form another graph, denoted by $G_3 \in \mathcal{F}_{pq}$, in such a way that we shift an edge/loop (which does not belong to $G_1 \cap G_2$ but belongs either to G_1 or to G_2) to a common "area" represented by $G_1 \cap G_2$. That is, we have constructed a graph $G_3 \in \mathcal{F}_{pq}$ such that $d(G_1, G_3) = d(G_2, G_3) = 2$. In other words, in the family \mathcal{F}_{pq} there should exist a path from G_1 to G_2 via the graph G_3 . As a by-product of these considerations is the following theorem.

THEOREM 3.4

The necessary and sufficient condition for a triple of graphs $G_1, G_2, G_3 \in \mathcal{F}_{pq}$ that the relation

$$d(G_1, G_2) + d(G_2, G_3) = d(G_1, G_3) \quad (3.12)$$

is fulfilled is that the graph G_2 lies on a minimal path connecting the graphs G_1 and G_3 in \mathcal{G}_{pq}^{CD} .

This theorem can be understood as an appendage to the triangle inequality (3.2c) in theorem 3.1.

3.1. REACTION GRAPH

The graph-theoretical approach is able to formulate a simple theory for describing chemical reactions [17]. Let us consider two graphs $G_1 = (V_1, E_1, L_1)$ and $G_2 = (V_2, E_2, L_2)$ from the family \mathcal{F}_{pq} . We introduce a nonsymmetric relation between these graphs called the *chemical transformation*,

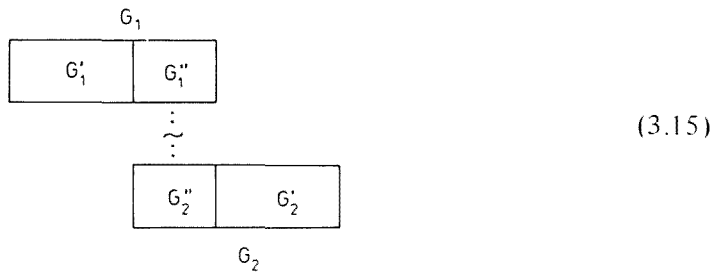
$$G_1 \Rightarrow G_2, \tag{3.13}$$

where $G_1(G_2)$ is called the *educt (product)* graph. Let $G_1 \cap G_2$ be a maximal common subgraph of G_1 and G_2 . The graphs may then be expressed as a "union" of two edge/loop disjoint subgraphs,

$$G_1 = G_1' \cup G_1'' \tag{3.14a}$$

$$G_2 = G_2' \cup G_2'' \tag{3.14b}$$

where the subgraphs $G_1'' \subseteq G_1$ and $G_2'' \subseteq G_2$ are isomorphic to $G_1 \cap G_2$, i.e. $G_1'' \approx G_2'' \approx G_1 \cap G_2$. Schematically,



We say that the transformation (3.13) changes the subgraph G_1' into the subgraph G_2' , whereas the subgraphs G_1'' and G_2'' isomorphic to $G_1 \cap G_2$ remain intact. Hence, the transformation (3.13) can be simplified as follows:

$$G_1' \Rightarrow G_2', \tag{3.16}$$

where we have omitted intact subgraphs of G_1 and G_2 . Let us consider a graph $\tilde{G}_2' = (\tilde{V}_2' = V_1', \tilde{E}_2', \tilde{L}_2')$, constructed over the same vertex set as the graph $G_1' = (V_1', E_1', L_1')$ and, moreover, isomorphic to G_2' , i.e. $\tilde{G}_2' \approx G_2'$.

DEFINITION 3.2

The *reaction graph* [17] G_R assigned to the transformation (3.16) is an ordered 5-tuple:

$$G_R = (V_R = V_1', E_R, L_R, \psi, \{-1, 1\}), \quad (3.17a)$$

$$E_R = E_1' \cup \tilde{E}_2', \quad (3.17b)$$

$$L_R = L_1' \cup \tilde{L}_2'. \quad (3.17c)$$

The mapping $\psi: E_R \cup L_R \rightarrow \{-1, 1\}$ evaluates the edges and loops by integers ± 1 ,

$$\psi(e) = \begin{cases} -1 & (\text{for } e \in E_1') \\ +1 & (\text{for } e \in \tilde{E}_2') \end{cases} \quad (3.18)$$

$$\psi(l) = \begin{cases} -1 & (\text{for } l \in L_1') \\ +1 & (\text{for } l \in \tilde{L}_2') \end{cases}. \quad (3.19)$$

The notion of the reaction graph (assigned to a fixed transformation $G_1' \Rightarrow G_2'$) enables us to "algebraicize" the transformation as follows:

$$G_1' + G_R = \tilde{G}_2' \approx G_2'. \quad (3.20)$$

The binary operation '+' is interpreted in the framework of a set-theoretical formalism as the symmetric difference [17] $A + B = (A \setminus B) \cup (B \setminus A)$,

$$A + B = \{x; (x \in A \wedge x \notin B) \vee (x \notin A \wedge x \in B)\}. \quad (3.21)$$

Applying this operation, the edge and loop sets of \tilde{G}_2' are

$$\tilde{E}_2' = E_1' + E_R, \quad \tilde{L}_2' = L_1' + L_R. \quad (3.22)$$

In the reaction graph G_R , those edges/loops that are cancelled (formed) are evaluated by -1 ($+1$).

DEFINITION 3.3

Two reaction graphs $G_R = (V_R, E_R, L_R, \psi, \{-1, 1\})$ and $G_R' = (V_R', E_R', L_R', \psi', \{-1, 1\})$ are *isomorphic* iff there exists such a one-to-one mapping $\omega: V_R \rightarrow V_R'$ that induces two one-to-one mappings $\chi: E_R \rightarrow E_R'$ and $\chi': L_R \rightarrow L_R'$ that conserve the incidences of edges and loops and their evaluation by ± 1 .

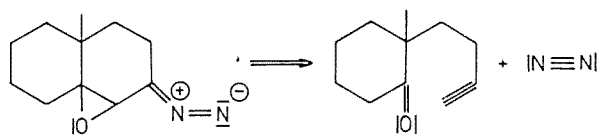
Let us consider two pairs G_1, G_2 and G_3, G_4 of graphs, both taken from the same family \mathcal{F}_{pq} of isomeric graphs. The graphs are related by two chemical transformations

$$G_1 \Rightarrow G_2, \quad G_3 \Rightarrow G_4 . \quad (3.23)$$

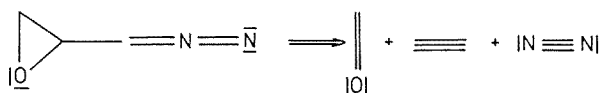
Applying the above theory, we assign to these transformations the reaction graphs G_R and G'_R , respectively. If the reaction graphs are isomorphic, then we say that the chemical transformations (3.23) are described by the same reaction pattern uniquely represented (up to the isomorphism) by a reaction graph. This property substantially increases the effectiveness of reaction graphs as a tool for classifying chemical reactions in organic chemistry. In particular, thousands of generic and named organic reactions are represented by just a few dozen reaction graphs.

EXAMPLE 3.6

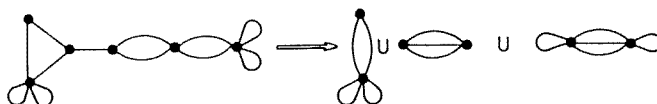
The Eschenmoser fragmentation [34] of 1,9-epoxy-2-diazo-10-methyl [4,4,0] bicyclodecane is



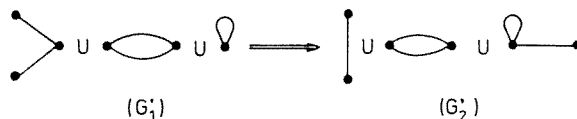
Removing the intact "environment", we get the basic skeleton of the fragmentation



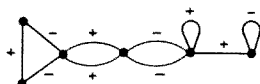
graphically represented by



Next, we remove untouched edges and loops,



Unifying the two sides of this pattern in a reaction graph, we arrive at a graph-theoretical representation of the Eschenmoser fragmentation



The reaction graph G_R assigned to the chemical transformation (3.13) is simply related to the chemical distance between its entries G_1 and G_2 .

THEOREM 3.5

The chemical distance between two isomeric graphs that are related by a chemical transformation $G_1 \Rightarrow G_2$ with assigned reaction graph G_R is determined by:

$$d(G_1, G_2) = |E_R| + |L_R|. \quad (3.24)$$

This theorem immediately follows from definition 3.2 of the reaction graph and, moreover, it has a very interesting chemical interpretation. One of the first heuristic rules of organic chemistry is the so-called *principle of minimal structural change* [35,36]. This states that in the course of a chemical transformation $G_1 \Rightarrow G_2$, the minimal number of edges/loops is created and/or annihilated in the educt graph G_1 to form the product graph G_2 . It implies that a reaction graph G_R assigned to the chemical transformation should be minimal with respect to the number of its edges and loops.

$$|E_R| + |L_R| = \min. \quad (3.25)$$

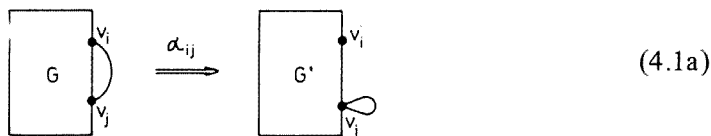
This requirement is automatically satisfied due to the definition of the reaction graph based on the notion of a maximal common subgraph. Ugi et al. [8,21] summarized the above property at the *principle of minimal chemical distance* as a very effective heuristic rule for the construction of reaction matrices.

4. Reaction distance

In the preceding section we have constructed the reaction graph G_R for a chemical transformation $G_1 \Rightarrow G_2$ on the basis of the maximal common subgraph $G_1 \cap G_2$. There exists an alternative possibility [31,32] for constructing this reaction graph, namely, to require that the chemical transformation is composed of a minimal number of the so-called elementary chemical transformations (or simply, elementary transformations), the number of which determines a reaction distance between the graphs G_1 and G_2 , and this distance induces a new metric for a fixed family of isomeric graphs.

Let us study the graph $G = (V, E, L)$ taken from the family \mathcal{F}_{pq} . We shall introduce two types of elementary transformations of the educt graph G into a product graph $G' = (V, E', L')$ with the same correspondence to the family \mathcal{F}_{pq} :

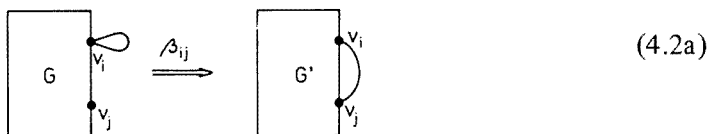
(1) The graph G contains the edge $[v_i, v_j]$. The *elementary chemical transformation* α_{ij} "dissociates" the edge $[v_i, v_j]$ on a loop $[v_i, v_j]$, schematically



The resulting product graph G' , isomorphic to a graph from the family \mathcal{F}_{pq} , has edge and loop sets determined as follows:

$$E' = E - \{[v_i, v_j]\}, \quad L' = L \cup \{[v_i, v_i]\}. \quad (4.1b)$$

(2) The graph G contains the loop $[v_i, v_i]$. The *elementary chemical transformation* β_{ij} "associates" the loop $[v_i, v_i]$ on an edge $[v_i, v_j]$, schematically

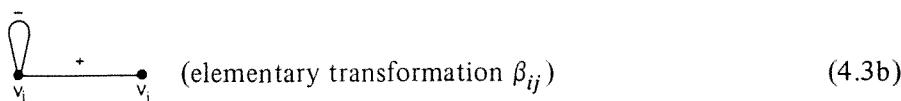
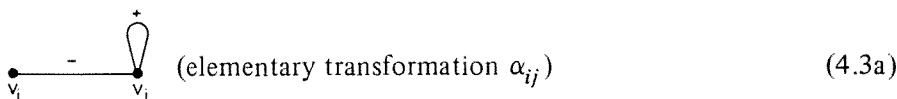


The resulting product graph G' , isomorphic to a graph from the family \mathcal{F}_{pq} , has edge and loop sets determined as follows:

$$E' = E \cup \{[v_i, v_j]\}, \quad L' = L - \{[v_i, v_i]\}. \quad (4.2b)$$

The elementary transformation β_{ij} may be formally treated as a retrotransformation with respect to the elementary transformation α_{ji} .

The two types of elementary transformations introduced above are simply visualized by the following reaction graphs:



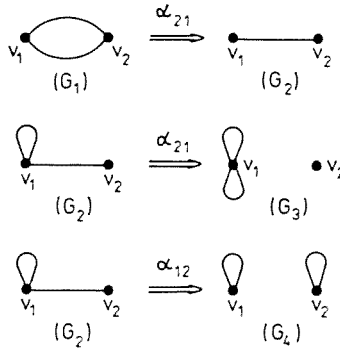
Hence, the chemical distance between G and G' for an arbitrary elementary transformation is $d(G, G') = 2$. A reverse property is not obeyed; $d(G, G') = 2$ does not imply that there exists an elementary transformation $\xi = \alpha, \beta$ such that $G \xrightarrow{\xi} G'$.

In a similar way as was done in section 3, we construct for a fixed family \mathcal{F}_{pq} of isomeric graphs the so-called *graph of reaction distances*, denoted by \mathcal{G}_{pq}^{RD} . Its vertex set is again identified with the family \mathcal{F}_{pq} . Two distinct vertices (corresponding

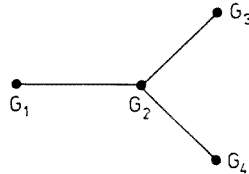
to nonisomorphic graphs $G, G' \in \mathcal{F}_{pq}$) are connected by an edge if there exists an elementary transformation $\xi = \alpha, \beta$ such that $G \xrightarrow{\xi} G'$.

EXAMPLE 4.1

Let us study the graphs from $\mathcal{F}_{2,2}$ (cf. example 2.3). For the following pairs of graphs there exist elementary transformations such that a member of the given pair is transformed into the other member of the same pair.



For the other remaining possible pairs of graphs from $\mathcal{F}_{2,2}$ there do not exist such elementary transformations. The graph of reaction distances constructed over the family $\mathcal{F}_{2,2}$ is



DEFINITION 4.1

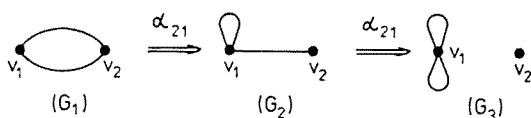
The reaction distance between two graphs $G, G' \in \mathcal{F}_{pq}$, denoted by $D(G, G')$, is equal to the graph distance between them in the graph \mathcal{G}_{pq}^{RD} .

EXAMPLE 4.2

The matrix $D = (D(G_i, G_j))$ of the reaction distance between graphs from $\mathcal{F}_{2,2}$ can be simply constructed from the $\mathcal{G}_{2,2}^{RD}$ presented in example 4.1.

$$D = \begin{pmatrix} 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 1 \\ 2 & 1 & 0 & 2 \\ 2 & 1 & 2 & 0 \end{pmatrix}$$

The chemical transformation $G_1 \Rightarrow G_3$ may be decomposed into two elementary transformations with an "intermediate" corresponding to the graph G_2 ,



This means that the reaction distance between G_1 and G_3 is $D(G_1, G_3) = 2$, in accordance with the graph $\mathcal{G}_{2,2}^{\text{RD}}$ presented in example 4.1. The graph distance between vertices 1 and 3 is equal to 2.

THEOREM 4.1

The reaction distance $D(G_1, G_2)$ is a *metric* and the following three properties are satisfied:

(i) Positive semidefiniteness

$$D(G_1, G_2) \geq 0 \quad (= 0 \text{ only for } G_1 \sim G_2). \quad (4.4a)$$

(ii) Symmetry

$$D(G_1, G_2) = D(G_2, G_1). \quad (4.4b)$$

(iii) Triangle inequality

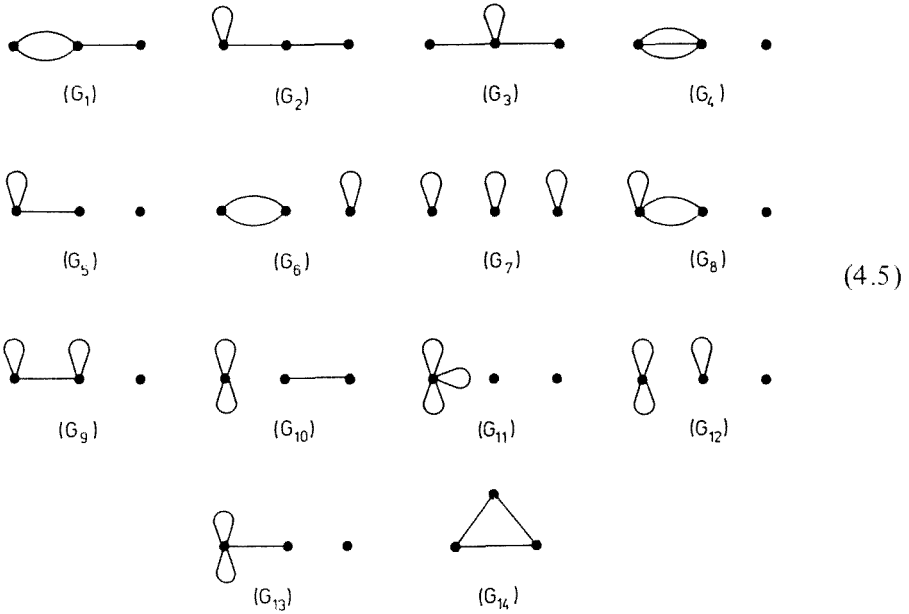
$$D(G_1, G_2) + D(G_2, G_3) \geq D(G_1, G_3). \quad (4.4c)$$

The equality holds iff the graph G_2 lies on the minimal path from G_1 to G_3 in $\mathcal{G}_{pq}^{\text{RD}}$.

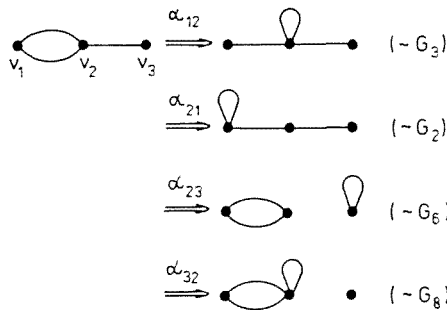
All these properties immediately follow from the definition of reaction distance as a graph distance in the graph $\mathcal{G}_{pq}^{\text{RD}}$.

4.1. ILLUSTRATIVE EXAMPLE

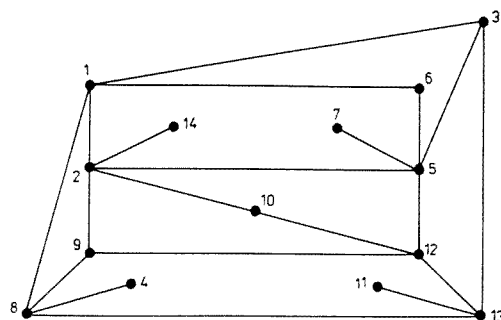
The above theory will be illustrated for the family $\mathcal{F}_{3,3}$ composed of fourteen graphs:



Let us consider, for instance, the graph G_1 . We apply all admissible elementary transformations,

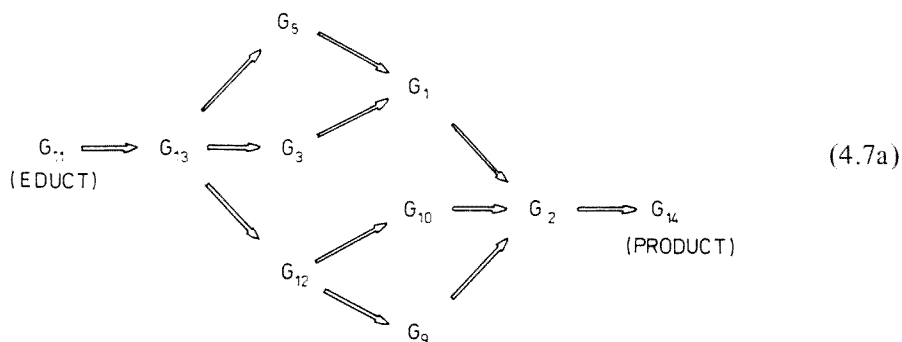


This means that in the graph $\mathcal{G}_{3,3}^{RD}$, the vertex v_1 (assigned to G_1) is adjacent to the vertices $v_2, v_3, v_6,$ and v_8 . Applying a similar procedure for all graphs from the family $\mathcal{F}_{3,3}$, we arrive at the following graph of reaction distances



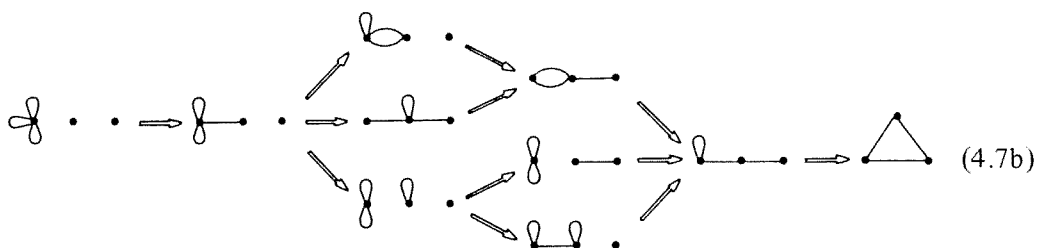
(4.6)

From this graph one may simply construct reaction distances between all graphs from $\bar{\mathcal{F}}_{3,3}$, e.g. $D(G_{11}, G_{14}) = 5$. Let us study the chemical transformation $G_{11} \Rightarrow G_{14}$. In accordance with the graph $\mathcal{G}_{3,3}^{RD}$, we obtain the following pattern:



(4.7a)

It will be much more instructive if the symbols are replaced by their diagrammatic representations:



(4.7b)

The decomposition of $G_{11} \Rightarrow G_{14}$ into five elementary transformations is

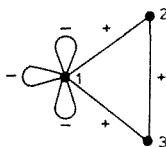
$$G_{11} \Rightarrow G_{13} \Rightarrow G_8 \Rightarrow G_1 \Rightarrow G_2 \Rightarrow G_{14}$$

$$G_{11} \Rightarrow G_{13} \Rightarrow G_3 \Rightarrow G_1 \Rightarrow G_2 \Rightarrow G_{14}$$

$$G_{11} \Rightarrow G_{13} \Rightarrow G_{12} \Rightarrow G_{10} \Rightarrow G_2 \Rightarrow G_{14}$$

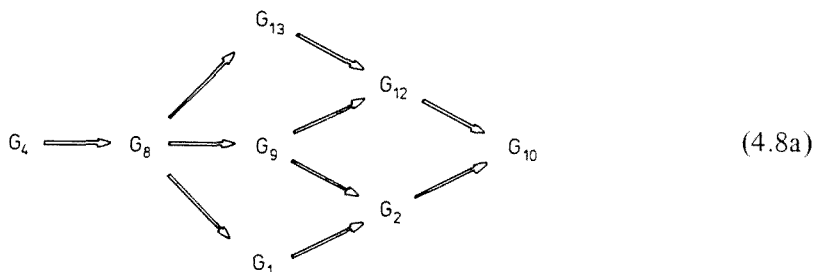
$$G_{11} \Rightarrow G_{13} \Rightarrow G_{12} \Rightarrow G_9 \Rightarrow G_2 \Rightarrow G_{14}$$

Applying chemical terminology, the transformation $G_{11} \Rightarrow G_{14}$ has four different mechanisms manifested by the sequences of elementary transformations α and β . The reaction graph corresponding to all four mechanisms is



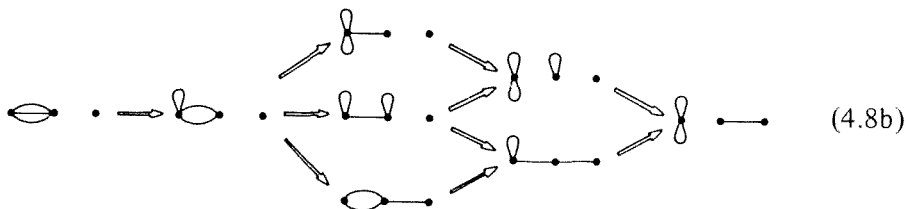
The same reaction graph is obtained on the basis of the maximal common subgraph.

We can analyze the transformation $G_4 \Rightarrow G_{10}$ in the same way. The reaction distance is $D(G_4, G_{10}) = 4$ and the assigned pattern of elementary transformation is



(4.8a)

diagrammatically



(4.8b)

The possible mechanisms for $G_4 \Rightarrow G_{10}$ are

$$G_4 \Rightarrow G_8 \Rightarrow G_{13} \Rightarrow G_{12} \Rightarrow G_{10}$$

$$G_4 \Rightarrow G_8 \Rightarrow G_9 \Rightarrow G_{12} \Rightarrow G_{10}$$

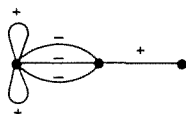
$$G_4 \Rightarrow G_8 \Rightarrow G_9 \Rightarrow G_2 \Rightarrow G_{10}$$

$$G_4 \Rightarrow G_8 \Rightarrow G_1 \Rightarrow G_2 \Rightarrow G_{10}$$

The reaction graph of $G_4 \Rightarrow G_{10}$, constructed on the basis of the maximal common subgraph $G_4 \cap G_{10}$, is



with a chemical distance $d(G_4, G_{10}) = 4$. The reaction graph constructed on the basis of the above mechanisms is



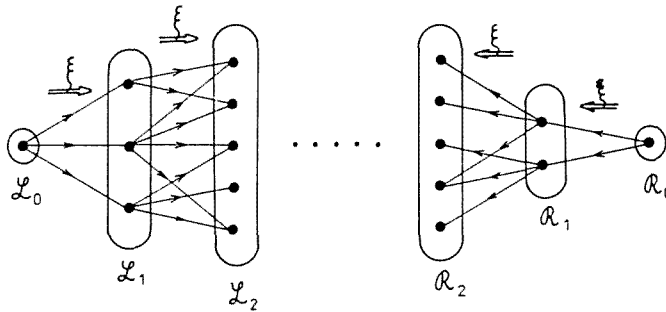
which does not correspond to the mentioned chemical distance $d(G_4, G_{10}) = 4$. We emphasize that the reaction graph constructed on the basis of the maximal common subgraph is decomposed into six elementary transformations, the result of which again is not compatible with the reaction distance $D(G_4, G_{10}) = 4$. Summarizing these observations, the chemical and reaction distances are not "isometric", therefore they may provide different reaction graphs.

4.2. BILATERAL APPROACH FOR EVALUATION OF REACTION DISTANCE

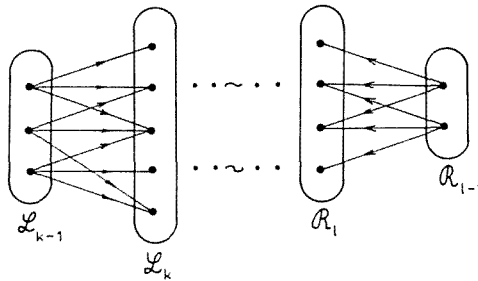
The chemical distance between two isomeric graphs was defined with the help of their maximal common subgraph. This definition makes it possible, in principle, to evaluate the chemical distances for an arbitrary pair of isomeric graphs. The reaction distance is defined as the graph distance in the graph of reaction distances. Therefore, in order to evaluate the reaction distance we have to know in advance the graph of reaction distances for a given family of isomeric graphs. Unfortunately, its explicit construction is possible only for relatively small families of \mathcal{F}_{pq} , i.e. for small values of the integers p and q (e.g. for $p, q \geq 10$, a construction of the graph $\mathcal{G}_{pq}^{\text{RD}}$ is a hopeless task). Therefore, it is very important to have an approach which does not require a knowledge of the whole graph of reaction distances but only a part of it, i.e. the subgraph which contains all the minimal paths for a preselected pair of graphs from \mathcal{F}_{pq} . Recently, Ugi and Fontain [37] suggested the so-called *bilateral approach* for the construction of all possible minimal paths between a fixed pair of graphs. We shall generalize this approach for the evaluation of reaction distances.

Let $G_1 = (V_1, E_1, L_1)$ and $G_2 = (V_2, E_2, L_2)$ be a pair of graphs from the family \mathcal{F}_{pq} . If these graphs are isomorphic, then their reaction distance is equal to zero, $D(G_1, G_2) = 0$ (cf. eq. (4.4a)). Therefore, we shall assume that the graphs G_1 and G_2 are nonisomorphic. Let us construct the left and right sets of graphs $\mathcal{L}_0 = \{G_1\}$ and $\mathcal{R}_0 = \{G_2\}$. The graph $G_1 \in \mathcal{L}_0$ is transformed by all admissible elementary transformations $\xi = \alpha, \beta$, and a new left set \mathcal{L}_1 is formed from all the nonisomorphic resulting graphs produced. In the k th step (if $k \geq 2$), the left set \mathcal{L}_k is formed only

from those graphs that are nonisomorphic with the graphs from \mathcal{L}_{k-2} . In a completely similar way, the right sets $\mathcal{R}_0, \mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_l$ are formed. Schematically



Let us assume that we have constructed the sets \mathcal{L}_k and \mathcal{R}_l (where either $k = l$ or $l \pm 1$). If these sets contain graphs that are isomorphic, formally $\mathcal{L}_k \cap \mathcal{R}_l \neq \emptyset$, schematically



then the reaction distance is $d(G_1, G_2) = k + l$. If this is not the case, we continue the process of construction of new left and right sets of graphs.

This bilateral approach for the evaluation of the reaction distance may be summarized in the form of the following algorithm.

ALGORITHM 4.1

Step 1. (Initialization). Input of graphs G_1 and G_2 ;

$$k := l := 0; \mathcal{L}_0 := \{G_1\}; \mathcal{R}_0 := \{G_2\};$$

Step 2. $k := k + 1$;

$\mathcal{L}_k := \{\text{nonisomorphic graphs constructed by elementary transformations of the graphs from } \mathcal{L}_{k-1}\}$;

if $k \geq 2$, then $\mathcal{L}_k := \mathcal{L}_k - \{\text{graphs isomorphic with graphs from } \mathcal{L}_{k-2}\}$;

if $\mathcal{L}_k \cap \mathcal{R}_l \neq \emptyset$, then go to step 4;

Step 3. $l = l + 1$;

$\mathcal{R}_l := \{\text{nonisomorphic graphs constructed by elementary transformations of the graphs from } \mathcal{R}_{l-1} \}$;
 if $l \geq 2$, then $\mathcal{R}_l := \mathcal{R}_l - \{\text{graphs isomorphic with graphs from } \mathcal{R}_{l-2}\}$;
 if $\mathcal{L}_k \cap \mathcal{R}_l \neq \emptyset$, then go to step 4, else go to step 2;

Step 4. Reaction distance $:= k + l$;

Step 5. End of algorithm.

As a by-product of this bilateral approach for the evaluation of reaction distances between isomeric graphs G_1 and G_2 , is the possibility to construct from the sets $\mathcal{L}_0, \mathcal{L}_1, \dots, \mathcal{L}_k$ and $\mathcal{R}_0, \mathcal{R}_1, \dots, \mathcal{R}_l$ all possible mechanisms of the chemical transformation $G_1 \Rightarrow G_2$ (or retrotransformation $G_2 \Rightarrow G_1$).

EXAMPLE 4.3

Apply the algorithm 4.1 for the evaluation of reaction distance between graphs G_4 and G_{10} from example 4.1.

Step 1. $k := 0, l := 0, \mathcal{L}_0 := \{G_4\}, \mathcal{R}_0 := \{G_{10}\}$.

Step 2. $k := 1, \mathcal{L}_1 := \{G_8\}, \mathcal{L}_1 \cap \mathcal{R}_0 = \emptyset$.

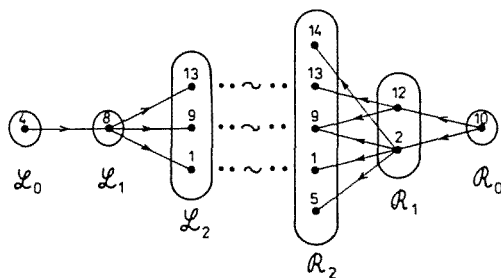
Step 3. $l := 1, \mathcal{R}_1 := \{G_2, G_{12}\}, \mathcal{L}_1 \cap \mathcal{R}_1 = \emptyset$.

Step 2. $k := 2, \mathcal{L}_2 := \{G_1, G_9, G_{13}\}, \mathcal{L}_2 \cap \mathcal{R}_1 = \emptyset$.

Step 3. $l := 2, \mathcal{R}_2 := \{G_1, G_5, G_9, G_{13}, G_{14}\}, \mathcal{L}_2 \cap \mathcal{R}_2 = \{G_1, G_9, G_{13}\} \neq \emptyset$.

Step 4. $D(G_4, G_{10}) := 2 + 2$.

Schematically, the above succession of steps can be illustrated by



From this scheme one can simply construct all possible mechanisms of the transformations $G_4 \Rightarrow G_{10}$ (cf. eqs. (4.9a–d)).

4.3. CONSTRUCTION OF PREDECESSORS AND SUCCESSORS FOR A GRAPH

A simple modification of the bilateral approach for the evaluation of reaction distances provides the so-called *monolateral approach* for the construction of all possible successors or predecessors for a given graph. Let $G = (V, E, L) \in \mathcal{F}_{pq}$ be a fixed graph. Successive applications of admissible elementary transformations α and/or β give all possible graphs (successors) with unit reaction distance from the graph G . From the first set of graphs (successors) we construct the next graphs, which are classified as successors of successors; their reaction distance from the graph G is 2. This recurrent process is continued until the prescribed reaction distance between G and its constructed successors is achieved. If we change the elementary transformations used to the corresponding retrotransformations, then the above procedure gives all predecessors of the given graph G with reaction distances bounded from above by the prescribed maximal reaction distance.

Let us turn our attention to a precise formulation of the monolateral approach for the construction of successors of the fixed graph G . First, we construct the set $\mathcal{G}_0 = \{G_0\}$. Applying all admissible elementary transformations α and/or β , we construct from the graph G all its first successors; their nonisomorphic representatives form the set \mathcal{G}_1 . From the graphs of \mathcal{G}_1 , we construct by all admissible elementary transformations the next set \mathcal{G}_2 , where this set does not contain the graphs isomorphic with graphs from \mathcal{G}_0 . This procedure is continued until we achieve the set $\mathcal{G}_{k_{\max}}$, where k_{\max} is the prescribed maximal reaction distance. The graphs from \mathcal{G}_i have the reaction distance i from the starting graph G .

ALGORITHM 4.2

Step 1. (Initialization). Input of graph G , k_{\max} ;

$$k := 0; \mathcal{G} := G;$$

Step 2. $k := k + 1$;

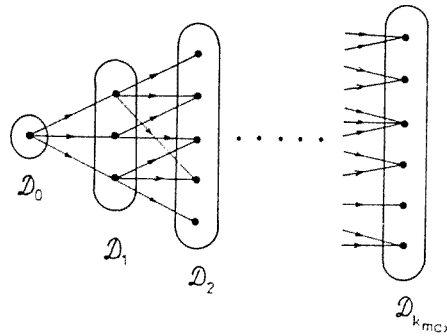
$\mathcal{G}_k := \{\text{nonisomorphic graphs constructed by elementary transformations of the graphs from } \mathcal{G}_{k-1}\}$;

if $k \geq 2$, then $\mathcal{G}_k := \mathcal{G}_k - \{\text{graphs isomorphic with graphs from } \mathcal{G}_{k-2}\}$;

Step 3. If $k < k_{\max}$, then go to step 2.

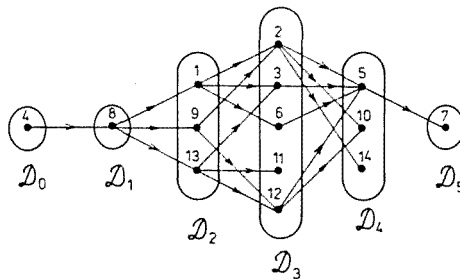
Step 4. End of algorithm.

The method of constructing all possible successors of the given graph G with the specified maximum reaction distance is schematically illustrated as follows:



EXAMPLE 4.4

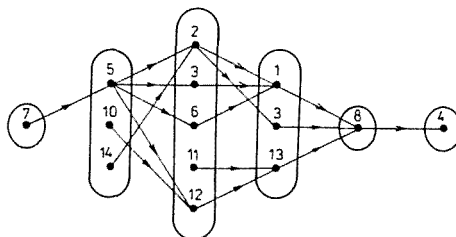
Construct all possible successors of the graph G_4 from (4.5) in the framework of the whole family $\mathcal{F}_{3,3}$. Successive applications of admissible elementary transformations to the graph G_4 provide



In this case, $k_{\max} = 5$ and we have constructed all possible successors of the graph G_4 in the family $\mathcal{F}_{3,3}$. For instance, the graph G_4 has a successor G_{11} , where $D(G_4, G_{11}) = 3$; decomposition of the transformation $G_4 \Rightarrow G_{11}$ into a sequence of elementary transformations gives

$$G_4 \Rightarrow G_8 \Rightarrow G_{13} \Rightarrow G_{11} .$$

Reversing the procedure, we get another scheme of all possible predecessors of G_4 in the whole family $\mathcal{F}_{3,3}$:



For instance, the graph G_4 may be "prepared" from the graph G_{11} by the following sequence of elementary transformations:

$$G_{11} \Rightarrow G_{13} \Rightarrow G_8 \Rightarrow G_4 .$$

5. Discussion

We have elaborated the graph-theoretical model of organic chemistry. This model is based on the concept of a molecular graph (a pseudomultigraph with vertices evaluated by the atomic symbols). Two different distances between the graphs were defined; both these distances induce a metric over the family \mathcal{F}_{pq} of isomeric graphs. First, the chemical distance based on the maximal common subgraph is closely related to Ugi et al.'s [8,9] matrix model of organic chemistry and their principle of minimal chemical distance. The resulting reaction graph assigned to the chemical transformation $G \Rightarrow G'$ is composed of a minimal number of edges and/or loops. Second, the reaction distance provides a very efficient method for the construction of mechanisms of chemical transformations. Unfortunately, these metrics are not "isometric", and therefore the reaction graphs constructed by making use of them may be, in general, different. The suggested graph-theoretical model of organic chemistry gives a very simple and transparent theoretical tool for the formal simulation of the organic chemistry in which all irrelevant details and exclusions are suppressed at minimum. It offers formal notions and concepts for an unambiguous description of computer algorithms for organic synthesis design based on the mechanisms of chemical reactions (the so-called second-generation programs, cf. ref. [8]).

A deep gap between our formal look at chemical reactions and their physical content based on the topology of potential surfaces may be abridged, likely, by an approach of Mezey [38,39]. He started, as was mentioned, from the potential surface of a molecular system (in general, this is the same for all isomeric molecules). The topology of the potential surface determines (its minima, maxima, and saddle points) the so-called reaction network. This concept is closely related to our graph of chemical and/or reaction distances, and determines possible reaction products (intermediates) that are neighbouring the preselected educt molecular structure on the potential surface.

Acknowledgements

We wish to express our thanks to our colleagues and friends Professor Milan Kratochvíl, Dr. Jaroslav Koča, Dr. Luděk Matyska, and Dr. Vladimír Baláž for their encouragement and valuable comments on different subjects of the mathematical model of organic chemistry.

References

- [1] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1969).
- [2] A.T. Balaban, ed., *Chemical Applications of Graph Theory* (Academic Press, London, 1976).
- [3] N. Trinajstić, *Chemical Graph Theory*, Vols. 1 and 2 (CRC Press, Boca Raton, FL, 1983).
- [4] R.B. King, ed., *Chemical Applications of Topology and Graph Theory* (Elsevier, Amsterdam, 1983).
- [5] I. Gutman and O. Polansky, *Mathematical Concepts in Organic Chemistry* (Springer-Verlag, Berlin, 1986).
- [6] N. Trinajstić, ed., *Mathematics and Computational Concepts in Chemistry* (Ellis Harwood, Chichester, UK, 1986).
- [7] E.J. Corey, *Pure Appl. Chem.* 14(1967)19.
- [8] J. Dugundji and I. Ugi, *Top. Curr. Chem.* 39(1973)19.
- [9] I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum and W. Schubert, *Angew. Chem. Int. Ed. Engl.* 18(1979)111.
- [10] J.B. Hendrickson, *Top. Curr. Chem.* 62(1976)49.
- [11] J.B. Hendrickson, E. Braun-Keller and G.A. Toczko, *Tetrahedron (Suppl.)* 37(1981)319.
- [12] N.S. Zefirov and S.S. Tratch, *Zh. Org. Khim.* 11(1975)225, 1785; 12(1976)697.
- [13] N.S. Zefirov, *Acc. Chem. Res.* 20(1987)237.
- [14] J. Koča, M. Kratochvil, V. Kvasnička, L. Matyska and J. Pospíchal, *A Synthron Model of Organic Chemistry and Synthesis Design*, in: *Lecture Notes in Chemistry*, Vol. 51 (Springer-Verlag, Berlin, 1989).
- [15] V. Kvasnička, *Coll. Czech. Chem. Comm.* 48(1983)2097.
- [16] V. Kvasnička, *Coll. Czech. Chem. Comm.* 48(1983)2118.
- [17] V. Kvasnička, M. Kratochvil and J. Koča, *Coll. Czech. Chem. Comm.* 48(1983)2284.
- [18] V. Kvasnička, *Coll. Czech. Chem. Comm.* 49(1984)1090.
- [19] V. Baláž, J. Koča, V. Kvasnička and M. Sekanina, *Časopis Pěst. Mat.* 111(1986)431.
- [20] V. Baláž, V. Kvasnička and J. Pospíchal, *Časopis Pěst. Mat.*, in press.
- [21] C. Jochum, J. Gasteiger, I. Ugi and J. Dugundji, *Z. Naturforsch.* 37b(1982)1205.
- [22] M.A. Johnson, in: *Graph Theory and its Application to Algorithm and Computer Science*, ed. Y. Alavi, G. Chartrand, L. Lesniak and C. Wall (Wiley, New York, 1985) p. 457.
- [23] M.A. Johnson, *Czech. Math. J.* 37(1987)75.
- [24] M.F. Lynch and P. Willet, *J. Chem. Inf. Comput. Sci.* 18(1978)154.
- [25] D. Bawden, T.K. Dewon, F.T. Jackson, S.I. Wood, M.F. Lynch and P. Willet, *J. Chem. Inf. Comput. Sci.* 19(1979)90.
- [26] P. Willet, *J. Chem. Inf. Comput. Sci.* 20(1980)93.
- [27] J.J. McGregor and P. Willet, *J. Chem. Inf. Comput. Sci.* 21(1981)139.
- [28] M. Wochner, J. Brandt, A. von Scholley and I. Ugi, *Chimia* 42(1988)217.
- [29] I. Ugi and M. Wochner, *J. Mol. Struct. (Theochem)* 1965(1988)229.
- [30] E. Fontain, J. Bauer and I. Ugi, *Z. Naturforsch.* 42b(1987)889.
- [31] J. Koča, *Coll. Czech. Chem. Comm.* 53(1988)3119.
- [32] J. Koča, *J. Math. Chem.* 3(1989)73; 91.
- [33] J.J. McGregor, *Soft.-Pract. Exp.* 12(1982)23.
- [34] H.O. House, *Modern Synthetic Reactions* (Benjamin, Menlo Park, 1972).
- [35] W. Hückel, *Theoretische Grundlagen der Organischen Chemie*, Band I (Academic Verlag, Leipzig, 1952).
- [36] G.H. Wheland, *Advanced Organic Chemistry* (Wiley, New York, 1962).
- [37] E. Fontain, J. Bauer and I. Ugi, *Chem. Lett.* (1987)37.
- [38] P.G. Mezey, *Theor. Chim. Acta* 60(1982)409.
- [39] P.G. Mezey, *Int. J. Quant. Chem. Symp.* 18(1984)675.