Theor Chim Acta (1991) 79: 65–79

Theoretica Chimica Acta

Reaction and chemical distances and reaction graphs

Vladimír Kvasnička, Jiří Pospíchal, and Vladimír Baláž
Department of Mathematics, Slovak Technical University, 81237 Bratislava, Czechoslovakia

Received August 13, 1990; received in revised form and accepted October 22, 1990

Summary. The distance between molecular graphs $\mathbb{G}$ and $\mathbb{G}'$ is equal to the length of a minimum path connecting these molecular graphs in the so-called graph of distances. The graph of distances has vertices which are molecular graphs taken from the same family of isomeric graphs, and two molecular graphs are adjacent if there exists a prototype reaction graph which transforms one into the other. Distances may alternatively be determined by applying the concept of common supergraphs. In particular, the reaction and chemical distances between isomeric molecular graphs are studied. These distances allow us to simply incorporate the principle of minimum structural change, often used in mechanistic organic chemistry.

Key words: Reaction graphs – Reaction distance – Chemical distance – Graph theory

1. Introduction

In mathematical models [1–5] of organic chemistry the concept of chemical distance, initially introduced by Dugundji and Ugi [1, 6], is a concept that reflects to some extent the very important principle of minimum structural change [7, 8] successfully used in organic chemistry for more than a century [9]. Following this principle a chemical reaction (transformation) of an educt molecule into a product molecule is realized in such a way that a minimum number of chemical bonds (and/or lone electron pairs) is canceled and formed. In Dugundji and Ugi's model of constitutional chemistry [1] molecules are represented by symmetric matrices closely related to the so-called adjacency matrices of graphs [10]. Let educt and product molecules be represented by the molecular graphs $\mathbb{G}$ and $\mathbb{G}'$ [11], and moreover, let these molecular graphs be represented by adjacency matrices $A = (A_{ij})$ and $A' = (A'_{ij})$. Then the chemical distance is determined [1, 6] as a minimum value of L_1 norm of a difference of matrices A and A'.

$$CD(\mathbb{G}, \mathbb{G}') = \min_{P} \tfrac{1}{2} \sum_{i \leqslant j} |A_{ij} - A'_{p_i p_j}|,$$

where $P = (p_1, p_2, \ldots, p_n)$ represents a permutation of $(1, 2, \ldots, n)$. An optimal permutation P_0 corresponds to such a matching between atoms of educt and product molecules that the maximum common substructure is saved in the course of the chemical reaction. In other words, for two graphs $\mathbb{G}$ and $\mathbb{G}'$ with the same number of vertices, the chemical distance between them is the minimum cardinality of the symmetric difference of the edge sets of the 2 graphs with regard to all possible labellings. Thus, labellings of $\mathbb{G}$ and $\mathbb{G}'$ are sought such that the intersection of the edge sets has maximum cardinality. Then the distance between them is given by the number of edges in the two graphs which are not in this intersection (cf. [19–22] and Theorem 6 in this work).

Ugi et al. [12–15] (cf. also Refs. [16–18]) have demonstrated that the chemical distance is a very important formal concept in that it is able to quantify a similarity or dissimilarity of molecules from the standpoint of the already mentioned principle of minimum structural change.

An alternative look at the problem of chemical distance was made within the framework of the so-called graph-theoretical model of organic chemistry [5] independently by Johnson [19] and Baláž et al. [20–22]. These authors have defined the chemical distance by resorting to the concept of maximum common subgraph and also proved that this type of graph distance is a metric (in particular, a triangle inequality may be proved).

Recently, Koča [23] (cf. also Refs. [5, 18, 22]) introduced the notion of reaction distance defined as the minimum number of elementary mechanistic (dissociation and association) steps which are necessary to transform an educt onto a product. This is closely related to the classical Ingold [24] concept of decomposition of an overall chemical reaction into a sequence of elementary push-pull electron flow acts. The formally correct definition of reaction distance may be done by making use of the so-called graph of distances. The vertices of this graph of distances are molecular graphs and two vertices, i.e. molecular graphs, are joined by an edge if one may be transformed into the other by application of a dissociation or association act. The reaction distance between two molecular graphs is then determined as the minimum length of a path connecting them. In our recent communications [18, 22] we have studied an interplay between chemical and reaction distances. We have demonstrated that in many chemically meaningful cases these distances are closely related and offer alternative possibilities for construction of reaction networks with acceptance of the principle of minimum structural change (now reformulated in principles of minimum chemical or reaction distance).

The purpose of this communication is to demonstrate that chemical and reaction distances may be introduced simultaneously either via the concept of minimum common supergraph or via the concept of the graph of distances, and that both these approaches are equivalent. The presented, detailed insight to these distances offers theoretical tools for better understanding and interpretation of links between them. Formally, our theoretical approach helps us to introduce many different distances which lie between the chemical and reaction distances having, unfortunately, more graph-theoretical than practical meaning.

2. Basic concepts

For the purposes of this communication it will be worthwhile to define the concept of graph [10] $\mathbb{G}$ as a formal structure composed of a nonempty vertex set $V(\mathbb{G}) = \{v_1, v_2, \ldots\}$, edge set $E(\mathbb{G}) = \{e_1, e_2, \ldots\}$, and a mapping $\varphi : E(\mathbb{G}) \rightarrow \{0, \pm 1, \pm 2, \ldots\}$. An edge $e \in E(\mathbb{G})$ incident with two vertices $v_1, v_2 \in V(\mathbb{G})$ (if these vertices are identical, then the edge is called a loop) will be denoted $\{v_1, v_2\}$. We shall postulate that the graph $\mathbb{G}$ does not contain multiple edges. The mapping φ assigns to each edge $e \in E(\mathbb{G})$ a nonzero integer. This mapping is extended outside the edge set $E(\mathbb{G})$ such that each edge $e \notin E(\mathbb{G})$ has zero image. The absolute value of $\varphi(e)$ is called the multiplicity of the edge e; edges not belonging to the edge set $E(\mathbb{G})$ are of zero multiplicity. In summary, the graph $\mathbb{G}$ is determined as an ordered triple:

$$\mathbb{G} = (V, E, \varphi), \tag{1}$$

where $V = V(\mathbb{G})$ and $E = E(\mathbb{G})$.

Two types of graphs will be distinguished in our considerations:

(1) *Molecular graph* – its edges are evaluated by positive integers.

(2) *Reaction graph*, its edges are evaluated by positive and/or negative integers and the mapping φ is constrained by

$$\sum_{e \in E(\mathbb{G})} \varphi(e) = 0 \tag{2}$$

Similar notion of the so-called signed graph was first introduced by Harary [27].

The graphs $\mathbb{G}_1 = (V_1, E_1, \varphi_1)$ and $\mathbb{G}_2 = (V_2, E_2, \varphi_2)$ are called isomorphic ($\mathbb{G}_1 \approx \mathbb{G}_2$) iff there exists a 1-1 mapping (isomorphism):

$$\chi : V(\mathbb{G}_1) \rightarrow V(\mathbb{G}_2) \tag{3}$$

which preserves the adjacency of vertices and the evaluation of edges. For each pair of adjacent vertices v_1, $v_2 \in V(\mathbb{G}_1)$ the corresponding mapped pair $\chi(v_1)$, $\chi(v_2) \in V(\mathbb{G}_2)$ is also adjacent and the evaluations of mutually assigned edges are identical:

$$\{v_1, v_2\} \in E(\mathbb{G}_1) \Leftrightarrow \{\chi(v_1), \chi(v_2)\} \in E(\mathbb{G}_2) \tag{4a}$$

$$\varphi_1(\{v_1, v_2\}) = \varphi_2(\{\chi(v_1), \chi(v_2)\}) \tag{4b}$$

The cardinality of $\mathbb{G}$ is determined as the sum of edge multiplicities:

$$|\mathbb{G}| = \sum_{e \in E(\mathbb{G})} |\varphi(e)| \tag{5}$$

In chemistry, two compounds are isomers if they both contain the same number of atoms of each kind and the same number of valence electrons grouped at bonds and lone pairs. For graphs a corresponding but more crude concept is graph isomerism which does not specify kinds of vertices corresponding to different types of atoms. Two molecular graphs $\mathbb{G}_1$ and $\mathbb{G}_2$ are isomeric if they have the same number of vertices and the same cardinalities:

$$|V(\mathbb{G}_1)| = |V(\mathbb{G}_2)| \tag{6a}$$

$$|\mathbb{G}_1| = |\mathbb{G}_2| \tag{6b}$$

The isomerism represents a relation of equivalence (which is reflexive, symmetric, and transitive). Hence a universe of graphs may be divided into disjoint families of mutually isomeric graphs. A family of molecular isomeric graphs $\mathcal{F}_{pq}$ is determined by:

$$\mathscr{F}_{pq} = \{\mathbb{G} = (V, E, \varphi);\ |V| = p \text{ and } |\mathbb{G}| = q\} \tag{7}$$

We say that a molecular graph $\mathbb{G}$ belongs to family $\mathscr{F}_{pq}$, $\mathbb{G} \in \mathscr{F}_{pq}$, if this family contains a graph isomorphic to $\mathbb{G}$.

An inverse graph $\bar{\mathbb{G}}$ to the graph $\mathbb{G}$ has the same vertex and edge sets as the graph $\mathbb{G}$ but the mapping $\bar{\varphi} = -\varphi$. Two graphs $\mathbb{G}_1$ and $\mathbb{G}_2$ are called symmetric if there exists a graph $\mathbb{G}$ such that $\mathbb{G} \approx \mathbb{G}_1$ and $\bar{\mathbb{G}} \approx \mathbb{G}_2$. In some special cases it may happen that graph $\mathbb{G}$ is isomorphic to its inverse graph, then we say that the graph $\mathbb{G}$ is selfsymmetric.

Graph $\mathbb{G}'$ is called the subgraph of graph $\mathbb{G}$ ($\mathbb{G}' \subseteq \mathbb{G}$) if $V(\mathbb{G}') \subseteq V(\mathbb{G})$, $E(\mathbb{G}') \subseteq E(\mathbb{G})$, and:

$$\forall e \in E(\mathbb{G}') : 0 < \varphi'(e) \leqslant \varphi(e) \quad \text{or} \quad \varphi(e) \leqslant \varphi'(e) < 0 \tag{8}$$

Also we may call $\mathbb{G}$ a supergraph of the graph $\mathbb{G}'$, $\mathbb{G} \supseteq \mathbb{G}'$.

A common subgraph $\tilde{\mathbb{G}}$ of two graphs $\mathbb{G}_1$ and $\mathbb{G}_2$, denoted by $\mathbb{G}_1 \wedge \mathbb{G}_2$, is simultaneously isomorphic to the subgraphs $\mathbb{G}_1' \subseteq \mathbb{G}_1$ and $\mathbb{G}_2' \subseteq \mathbb{G}_2$, i.e. $\tilde{\mathbb{G}} \approx \mathbb{G}_1' \approx \mathbb{G}_2'$. A maximum common subgraph of $\mathbb{G}_1$ and $\mathbb{G}_2$, denoted $\mathbb{G}_1 \cap \mathbb{G}_2$, is a common subgraph with the highest possible cardinality:

$$|\mathbb{G}_1'| = |\mathbb{G}_2'| = |\mathbb{G}_1 \cap \mathbb{G}_2| = \max \tag{9}$$

A common supergraph of two graphs $\mathbb{G}_1$ and $\mathbb{G}_2$ consists of a graph $\tilde{\mathbb{G}}$ (called the supergraph and denoted by $\mathbb{G}_1 \vee \mathbb{G}_2$) such that it contains two subgraphs which are simultaneously isomorphic to the graphs $\mathbb{G}_1$ and $\mathbb{G}_2$, respectively, $\mathbb{G}_1 \approx \mathbb{G}_1' \subseteq \tilde{\mathbb{G}}$ and $\mathbb{G}_2 \approx \mathbb{G}_2' \subseteq \tilde{\mathbb{G}}$. A minimal common supergraph of $\mathbb{G}_1$ and $\mathbb{G}_2$, denoted by $\mathbb{G}_1 \cup \mathbb{G}_2$, is a common supergraph with the lowest possible cardinality:

$$|\mathbb{G}_1 \cup \mathbb{G}_2| = \min \tag{10}$$

We emphasize that concepts of common subgraph and supergraph may be simply generalized by induction for more than two graphs. Thus, for instance, the maximum common subgraph of a triple of graphs $\mathbb{G}_1$, $\mathbb{G}_2$, $\mathbb{G}_3$ will be denoted by $\mathbb{G}_1 \cap \mathbb{G}_2 \cap \mathbb{G}_3$, and be hence isomorphic to a common subgraph $\mathbb{G}_1 \wedge \mathbb{G}_2 \wedge \mathbb{G}_3$ with highest cardinality.

Example 1. Let us study two graphs $\mathbb{G}_1$ and $\mathbb{G}_2$ in Fig. 1. The maximum (minimum) common subgraph (supergraph) assigned to these graphs is displayed in the second row in Fig. 1. Displayed in the third row are a common subgraph and supergraph that are not extremum, i.e. maximum or minimum. □

Now let us focus our attention on how to specify a reaction graph $\mathbb{G}_R$ so that, when applied to a molecular graph $\mathbb{G}_1$, it produces a new graph $\mathbb{G}_2$. We shall postulate that the cardinality of $V(\mathbb{G}_R)$ is smaller than or equal to the cardinality of $V(\mathbb{G}_1)$, i.e. $|V(\mathbb{G}_R)| \leqslant |V(\mathbb{G}_1)|$. Since these vertex sets are, in general, different, we introduce a 1-1 mapping of $V(\mathbb{G}_R)$ onto a subset $\tilde{V}(\mathbb{G}_1) \subseteq V(\mathbb{G}_1)$,

$$\omega : V(\mathbb{G}_R) \to \tilde{V}(\mathbb{G}_1). \tag{11}$$

The mapping ω specifies the correspondence between vertices of the reaction graph $\mathbb{G}_R$ and vertices of a subgraph of $\mathbb{G}_1$ induced [10] by the subset $\tilde{V}(\mathbb{G}_1)$. In our forthcoming considerations, we shall have an inverse mapping

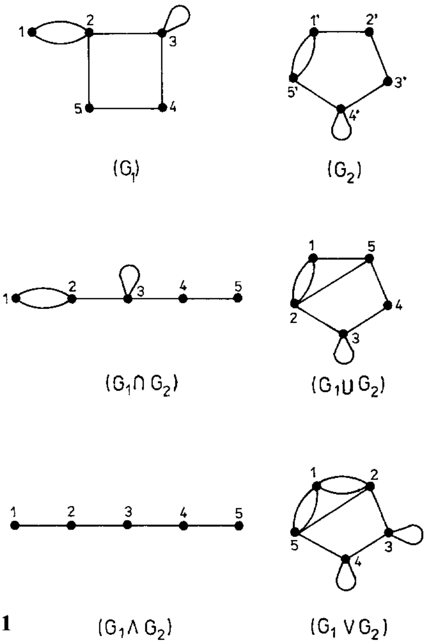


Fig. 1. The common subgraphs (supergraphs) of the graphs $\mathbb{G}_1$ and $\mathbb{G}_2$

$\omega^{-1}: \tilde{V}(\mathbb{G}_1) \to V(\mathbb{G}_R)$ extended outside of $\tilde{V}(\mathbb{G}_1)$ to the whole vertex set $V(\mathbb{G}_1)$, where $\omega^{-1}(v) = z$ for each $v \in V(\mathbb{G}_1)\backslash\tilde{V}(\mathbb{G}_1)$ and z is a "virtual" vertex, i.e. one that does not belong to $V(\mathbb{G}_R)$. The resulting graph $\mathbb{G}_2$ formed from the molecular graph $\mathbb{G}_1$ by the reaction graph $\mathbb{G}_R$ and the mapping ω is formally determined as an ordered triple:

$$(\mathbb{G}_1, \omega, \mathbb{G}_R) = \mathbb{G}_2 \tag{12}$$

Its vertex set $V(\mathbb{G}_2)$ is equal to the vertex set $V(\mathbb{G}_1)$; the edge set $E(\mathbb{G}_2)$ is determined through the mapping φ_2 as follows:

$$\varphi_2(\{v_1, v_2\}) = \varphi_1(\{v_1, v_2\}) + \varphi_R(\{\omega^{-1}(v_1), \omega^{-1}(v_2)\}) \tag{13}$$

Formally, the process of construction of $\mathbb{G}_2$ from $\mathbb{G}_1$ will be considered a transformation of $\mathbb{G}_1$ to $\mathbb{G}_2$:

$$\mathbb{G}_1 \Rightarrow \mathbb{G}_2 \tag{14}$$

specified by a reaction graph $\mathbb{G}_R$ and a mapping ω.

We call the transformation of Eq. (14) [or Eq. (12)] feasible if the mapping φ_2 is nonnegative (i.e. the resulting graph $\mathbb{G}_2$ is a molecular graph), $\varphi_2(\{v_1, v_2\}) \geqslant 0$. Otherwise, when for a pair of vertices $v_1, v_2 \in V(\mathbb{G}_1)$ we have $\varphi_2(\{v_1, v_2\}) < 0$, the reaction graph $\mathbb{G}_R$ specified by the mapping ω forms from the molecular graph $\mathbb{G}_1$ a new graph $\mathbb{G}_2$ which is not molecular. The edge set $E(\mathbb{G}_2)$ of the molecular graph $\mathbb{G}_2$ produced by a feasible transformation $(\mathbb{G}_1, \omega, \mathbb{G}_R)$ is determined by:

$$E(\mathbb{G}_2) = \{\{v_1, v_2\};\ \varphi_2(\{v_1, v_2\}) > 0\} \tag{15}$$

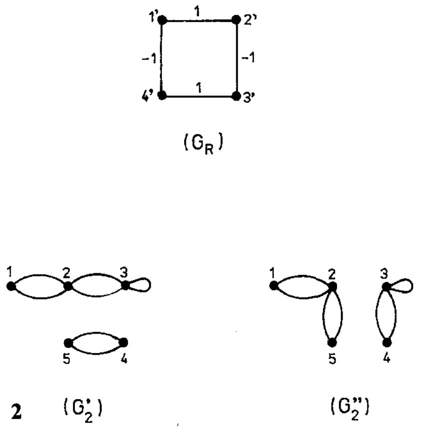


Fig. 2. Application of the reaction graph $\mathbb{G}_R$, specified by two different mappings ω_1' and ω_1'' in example 2, onto the molecular graph $\mathbb{G}_1$ in Fig. 1 provides the molecular graphs $\mathbb{G}_2'$ and $\mathbb{G}_2''$ given in the second row

Example 2. Let us study the molecular graph $\mathbb{G}_1$ in Fig. 1 and let the reaction graph $\mathbb{G}_R$ be specified by the graph shown in the first row in Fig. 2. The following two mappings will be used for the specification of $\mathbb{G}_R$:

$$\omega_1'(2) = 1', \quad \omega_1'(3) = 2', \quad \omega_1'(4) = 3', \quad \omega_1'(5) = 4'$$

$$\omega_1''(2) = 1', \quad \omega_1''(3) = 4', \quad \omega_1''(4) = 3', \quad \omega_1''(5) = 2'$$

The resulting molecular graphs are displayed in the second row in Fig. 2. □

If molecular graphs $\mathbb{G}_1$ and $\mathbb{G}_2$ are related by a feasible transformation of Eq. (14) and $\mathbb{G}_1$ belongs to the family $\mathscr{F}_{pq}$, then the produced graph $\mathbb{G}_2$ belongs also to the same family $\mathscr{F}_{pq}$ of isomeric graphs. This property implies that the transformation of Eq. (14) preserves the number of vertices as well as the cardinality of graphs as an immediate consequence of the definition of transformation and the property in Eq. (2) of reaction graphs. Hence, claiming that two molecular graphs are mutually related by a transformation we automatically assume that these graphs are also isomeric.

If two molecular graphs $\mathbb{G}_1, \mathbb{G}_2 \in \mathscr{F}_{pq}$ are related by $\mathbb{G}_1 \Rightarrow \mathbb{G}_2$, i.e. $(\mathbb{G}_1, \omega_1, \mathbb{G}_R) = \mathbb{G}_2$, then the inverse reaction graph $\bar{\mathbb{G}}_R$ performs a retrotransformation $\mathbb{G}_2 \Rightarrow \mathbb{G}_1$, where $(\mathbb{G}_2, \omega_2, \bar{\mathbb{G}}_R) = \mathbb{G}_1$. This property immediately follows from Eq. (2) required for reaction graphs. If the transformation $\mathbb{G}_1 \Rightarrow \mathbb{G}_2$ is realized by a reaction graph $\mathbb{G}_R$ (and mapping ω_1), then the retrotransformation $\mathbb{G}_2 \Rightarrow \mathbb{G}_1$ is carried out by an inverse reaction graph $\bar{\mathbb{G}}_R$ (and mapping ω_2).

Finally, let us introduce a few new concepts and notions which will be frequently used in our forthcoming considerations. Let $\mathscr{R}$ be a nonempty set composed of the so-called prototype reaction graphs which are pairwise nonisomorphic. Moreover, if $\mathbb{G}_R$ belongs to $\mathscr{R}$, then also a reaction graph $\mathbb{G}_R'$, isomorphic to $\bar{\mathbb{G}}_R$, belongs to $\mathscr{R}$. This requirement means that the set $\mathscr{R}$ is formed either of couples of symmetric reaction graphs or single selfsymmetric reaction graphs. It ensures that if transformation $\mathbb{G}_1 \Rightarrow \mathbb{G}_2$ is carried out by reaction graph $\mathbb{G}_R \in \mathscr{R}$, then retrotransformation $\mathbb{G}_2 \Rightarrow \mathbb{G}_1$ is also feasible and carried out by a reaction graph from $\mathscr{R}$, in particular, by a reaction graph isomorphic to the inverse reaction graph $\bar{\mathbb{G}}_R \in \mathscr{F}$. A graph $\mathbb{G}_{pq}(\mathscr{R})$, called the graph of distances, is determined in such a way that its vertex set is identified with the family $\mathcal{F}_{pq}$. Two distinct vertices (corresponding to a pair of nonisomorphic molecular graph $\mathbb{G}$, $\mathbb{G}' \in \mathcal{F}_{pq}$) are adjacent by an edge if there exists a reaction graph $\mathbb{G}_R \in \mathcal{R}$ (specified by a mapping ω) such that $(\mathbb{G}, \omega, \mathbb{G}_R) = \mathbb{G}'$, that is, the isomeric molecular graphs $\mathbb{G}$, $\mathbb{G}'$ may be related by the binary relations called transformation $\mathbb{G} \Rightarrow \mathbb{G}'$ and retrotransformation $\mathbb{G}' \Rightarrow \mathbb{G}$, carried out by the reaction graphs $\mathbb{G}_R$ and $\bar{\mathbb{G}}_R$, respectively.

In general, the graph of distances $\mathbb{G}_{pq}(\mathcal{R})$ may be disconnected, it contains two or more components. For a given component, its two distinct vertices (corresponding to a pair of nonisomorphic molecular graphs belonging to the family $\mathcal{F}_{pq}$) are connected by a path (composed of a sequence of edges). Let us denote by $d(\mathbb{G}, \mathbb{G}')$ the length of a minimal past between two vertices (corresponding to $\mathbb{G}, \mathbb{G}' \in \mathcal{F}_{pq}$) from the same component. This entity may be equal to zero iff the molecular graphs are isomorphic. The graph distance between molecular graphs belonging to the family $\mathcal{F}_{pq}$ is determined (with respect to the set $\mathcal{R}$ of prototype reaction graphs) as follows:

$$D(\mathbb{G}, \mathbb{G}') = \begin{cases} d(\mathbb{G}, \mathbb{G}'), & \text{(Graphs } \mathbb{G} \text{ and } \mathbb{G}' \text{ are from} \\ & \quad \text{the same component of } \mathbb{G}_{pq}(\mathcal{R})) \\ \infty & \text{(Graphs } \mathbb{G} \text{ and } \mathbb{G}' \text{ are from} \\ & \quad \text{different components of } \mathbb{G}_{pq}(\mathcal{R})) \end{cases} \tag{16}$$

We see that the concept of distance between pairs of molecular graphs is well determined only when these pairs belong to the same component of the distance graph $\mathbb{G}_{pq}(\mathcal{R})$. In the opposite case, i.e. the graphs belong to different components of $\mathbb{G}_{pq}(\mathcal{R})$, the distance $D(\mathbb{G}, \mathbb{G}')$ is formally set equal to "infinity".

It is intriguing to formulate necessary and sufficient conditions required for prototype reaction graphs of $\mathcal{R}$ such that the corresponding graph of distances $\mathbb{G}_{pq}(\mathcal{R})$ will be connected. A partial solution of this problem may be achieved in a form of sufficient conditions ensuring the connectivity of $\mathbb{G}_{pq}(\mathcal{R})$. Let us consider a pair $\mathbb{G}$, $\mathbb{G}'$ of graphs from $\mathcal{F}_{pq}$ such that these graphs only differ at two edges, or two loops or at one edge and one loop. This means that starting from a maximum common subgraph $\mathbb{G} \cap \mathbb{G}'$ we may form graphs which are isomorphic to $\mathbb{G}$ and $\mathbb{G}'$ by adding an edge or a loop, i.e. $\mathbb{G} \cap \mathbb{G}' + \{e\} \approx \mathbb{G}$ and $\mathbb{G} \cap \mathbb{G}' + \{e'\} \approx \mathbb{G}'$. More precisely, the graphs $\mathbb{G}$ and $\mathbb{G}'$ are related by $|\mathbb{G} \cap \mathbb{G}'| + 1 = q$.

Theorem 1. If for an arbitrary pair of graphs $\mathbb{G}, \mathbb{G}' \in \mathcal{F}_{pq}$ restricted by $|\mathbb{G} \cap \mathbb{G}'| + 1 = q$ a transformation $\mathbb{G} \Rightarrow \mathbb{G}'$ may be decomposed into a sequence of elementary transformations carried out by prototype reaction graphs from $\mathcal{R}$, then the corresponding graph of distances $\mathbb{G}_{pq}(\mathcal{R})$ is connected. □

Loosely speaking, the above theorem states that if an arbitrary edge/loop could be moved into another one by prototype reaction graphs from $\mathcal{R}$, then the graphs of distances should be connected. This conclusion is obvious, since an arbitrary graph may be transformed into another one by a process which moves an edge/loop into other positions and also changes an edge into a loop and vice versa.

Therefore, in order to avoid the above mentioned formal difficulties in our forthcoming considerations on different specific distances which will induce metrics, we shall always require that the set $\mathcal{R}$ of prototype reaction graphs is chosen in such a way that the corresponding graphs of distances $\mathbb{G}_{pq}(\mathcal{R})$ is connected. Then the distance $D(\mathbb{G}, \mathbb{G}')$ is a nonnegative integer, vanishing iff the molecular graphs $\mathbb{G}$, $\mathbb{G}'$ are isomorphic:

$$D(\mathbb{G}, \mathbb{G}') \geqslant 0 \quad (=0 \text{ iff } \mathbb{G} \approx \mathbb{G}') \tag{17}$$

Moreover, the graph distance is symmetric and the triangle inequality satisfied:

$$D(\mathbb{G}, \mathbb{G}') = D(\mathbb{G}', \mathbb{G}) \tag{18}$$

$$D(\mathbb{G}, \mathbb{G}'') + D(\mathbb{G}'', \mathbb{G}') \geqslant D(\mathbb{G}, \mathbb{G}') \tag{19}$$

where the sign of equality is satisfied iff the molecular graph $\mathbb{G}''$ lies on a shortest path connecting the terminal graphs $\mathbb{G}$ and $\mathbb{G}'$ in $\mathbb{G}_{pq}(\mathscr{R})$.

Theorem 2. If the set $\mathscr{R}$ of prototype reaction graphs is chosen in such a way that the graph of distances $\mathbb{G}_{pq}(\mathscr{R})$ is connected, then the graph distance $D(\mathbb{G}, \mathbb{G}')$ is a metric over the family $\mathscr{F}_{pq}$ of isomeric molecular graphs, i.e. $(\mathscr{F}_{pq}, D)$ is a finite metric space. □

3. Graph distances

At the end of the previous section we introduced the so-called graph distance between molecular graphs of the same family $\mathscr{F}_{pq}$, whereby the graph distance is related to a set $\mathscr{R}$ composed of prototype reaction graphs. We remember that these prototype reaction graphs are either pairwise symmetric or as singles selfsymmetric. The set $\mathscr{R}$ will be always chosen in such a way that the corresponding graph of distances $\mathbb{G}_{pq}(\mathscr{R})$ is connected, that is it may be applied for definitions of graph distances, see Eq. (21).

Theorem 3. Let $\mathscr{R}'$ be a set of prototype reaction graphs and let it be a subset of a "larger" set $\mathscr{R}$ of prototype reaction graphs, $\mathscr{R}' \subseteq \mathscr{R}$. The following three properties are satisfied:

(1) If the graph $\mathbb{G}_{pq}(\mathscr{R}')$ is connected, then the graph $\mathbb{G}_{pq}(\mathscr{R})$ is also connected.

(2) The graph $\mathbb{G}_{pq}(\mathscr{R}')$ is a subgraph of $\mathbb{G}_{pq}(\mathscr{R})$, $\mathbb{G}_{pq}(\mathscr{R}') \subseteq \mathbb{G}_{pq}(\mathscr{R})$.

(3) If the graph distances $D'(\mathbb{G}, \mathbb{G}')$ and $D(\mathbb{G}, \mathbb{G}')$ are determined over the graphs $\mathbb{G}_{pq}(\mathscr{R}')$ and $\mathbb{G}_{pq}(\mathscr{R})$, respectively, then:

$$0 \leqslant D(\mathbb{G}, \mathbb{G}') \leqslant D'(\mathbb{G}, \mathbb{G}'). \qquad \square \tag{20}$$

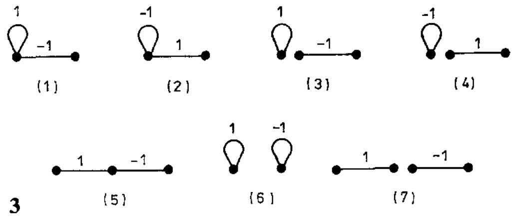
Proofs of these three statements are very simple. Since $\mathscr{R}' \subseteq \mathscr{R}$, assuming that $\mathbb{G}_{pq}(\mathscr{R}')$ is connected, then $\mathbb{G}_{pq}(\mathscr{R})$ should be also connected. The next two statements are also immediate consequences of $\mathscr{R}' \subseteq \mathscr{R}$. Each pair of graphs $\mathbb{G}$, $\mathbb{G}'$ that are joined in $\mathbb{G}_{pq}(\mathscr{R}')$ by a path must be also joined by a path with length shorter than or equal to a length of the former path.

Reaction distance. The set $\mathscr{R}$ of prototype reaction graphs is chosen as:

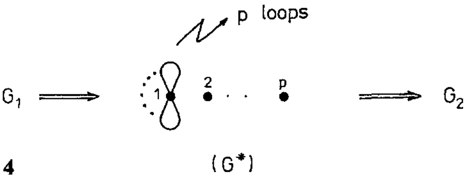
$$\mathscr{R} = \{\mathbb{G}_R^{(1)}, \mathbb{G}_R^{(2)}\} \tag{21}$$

where the reaction graphs $\mathbb{G}_R^{(1)}$ and $\mathbb{G}_R^{(2)}$ are isomorphic to the graphs denoted by 1 and 2 in Fig. 3. We see that this set $\mathscr{R}$ is composed of one pair of mutually symmetric graphs, $\mathbb{G}_R^{(1)} \approx \bar{\mathbb{G}}_R^{(2)}$.

Theorem 4. The graph $\mathbb{G}_{pq}(\mathscr{R})$ is connected. □



3



4

Fig. 3. All possible reaction graphs with cardinality equal to 2. The couples 1-2 and 3-4 correspond to pairs of mutually symmetric reaction graphs whereas the singles 5, 6, and 7 are selfsymmetric

Fig. 4. An illustrative example how to transform a molecular graph $\mathbb{G}_1$ onto another molecular graph $\mathbb{G}_2$, whereby $\mathbb{G}_1, \mathbb{G}_2 \in \mathcal{F}_{pq}$. The intermediate graph $\mathbb{G}^*$ is composed of p vertices and q loops incident with the same vertex

In order to prove this theorem we shall use a possibility outlined in Theorem 2. Let us select from the family $\mathcal{F}_{pq}$ a graph $\mathbb{G}^*$ composed of q loops incident with the same vertex, see Fig. 4. It is now easy to show that each graph $\mathbb{G}_1 \in \mathcal{F}_{pq}$ may be transformed by application of a finite number of reaction operators from the set $\mathcal{R}$ determined by Eq. (21) onto the graph $\mathbb{G}^*$. This also means that the graph $\mathbb{G}^*$ may be analogously (retro-) transformed onto another graph $\mathbb{G}_2 \in \mathcal{F}_{pq}$, formally (see Fig. 4):

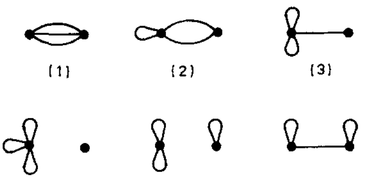
$$\mathbb{G}_1 \Rightarrow \mathbb{G}^* \Rightarrow \mathbb{G}_2.$$

Hence, we have proved that between a pair of arbitrary isomeric graphs there exists a path with terminal vertices $\mathbb{G}_1$ and $\mathbb{G}_2$ and going through the graph $\mathbb{G}^*$, that is the graph of distances $\mathbb{G}_{pq}(\mathcal{R})$ is connected, which was to be demonstrated.

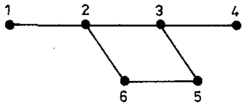
The distance $D(\mathbb{G}, \mathbb{G}')$ with the choice of Eq. (21) will be called the reaction distance and will be denoted by $RD(\mathbb{G}, \mathbb{G}')$. It was initially described by Koča [23] (see also Refs. [5, 18, 22]) as a proper graph-theoretic tool for treatment of chemical transformations (reactions) of an educt molecular graph into a product molecular graph. It allows us to decompose an "overall" transformation $\mathbb{G} \Rightarrow \mathbb{G}'$ onto a sequence of the so-called elementary transformations realized by the prototype reaction graph from Eq. (21):

$$\mathbb{G} \approx \mathbb{G}_0 \Rightarrow \mathbb{G}_1 \Rightarrow \mathbb{G}_2 \Rightarrow \mathbb{G}_3 \Rightarrow \cdots \Rightarrow \mathbb{G}_n \approx \mathbb{G}' \tag{22}$$

where the "intermediate" molecular graphs $\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_{n-1} \in \mathcal{F}_{pq}$ correspond to those vertices – graphs that form a shortest path connecting the vertices – graphs $\mathbb{G}$ and $\mathbb{G}'$ in the graph of distances $\mathbb{G}_{pq}(\mathcal{R})$. The length of sequence of Eq. (22) is equal to the reaction distance $RD(\mathbb{G}, \mathbb{G}')$, i.e. it is minimized from the standpoint of definition of RD. This is a manifestation of the well-known principle of minimum structural change often used by chemists when they are considering a mechanism (i.e. sequence of elementary chemical transformations) of an overall chemical transformation.



5



6

Fig. 5. All molecular graphs from the family of isomeric graphs $\mathcal{F}_{2,3}$

Fig. 6. Graph of distances determined over the family $\mathcal{F}_{2,3}$ and with respect to the set $\mathcal{R}$ of prototype reaction graphs in Eq. (21). The length of the shortest path between vertices – molecular graphs $\mathbb{G}, \mathbb{G}' \in \mathcal{F}_{2,3}$ is equal to the reaction distance $RD(\mathbb{G}, \mathbb{G}')$ between them

Example 3. Let us consider a family $\mathcal{F}_{2,3}$ of isomeric molecular graphs, see Fig. 5. The corresponding graph $\mathbb{G}_{2,3}(\mathcal{R})$, where the set $\mathcal{R}$ is determined by Eq. (21), is displayed in Fig. 6. Then the reaction distances between the graphs from $\mathcal{F}_{2,3}$ are listed in the form of the following matrix

$$RD = \begin{pmatrix} 0 & 1 & 2 & 3 & 3 & 2 \\ 1 & 0 & 1 & 2 & 2 & 1 \\ 2 & 1 & 0 & 1 & 1 & 2 \\ 3 & 2 & 1 & 0 & 2 & 3 \\ 3 & 2 & 1 & 2 & 0 & 1 \\ 2 & 1 & 2 & 3 & 1 & 0 \end{pmatrix}. \qquad \square$$

The sequence in Eq. (22) of the elementary transformations carried out by prototype reaction graphs from the set in Eq. (21) may be expressed recurrently as follows [see Eq. (12)]:

$$(\mathbb{G}_i, \omega_i, \mathbb{G}_R^{[i]}) = \mathbb{G}_{i+1}, \tag{23}$$

for $i = 0, 1, 2, \ldots, n-1$, where $\omega_0, \omega_1, \ldots, \omega_{n-1}$ are mappings (see Eq. (11)) specifying actions of reaction graphs $\mathbb{G}_R^{[1]}, \mathbb{G}_R^{[2]}, \ldots, \mathbb{G}_R^{[n-1]}$, respectively on the so-called predecessor molecular graphs. Applying successively the concept of common supergraph introduced in the previous section, we can construct a special type of common supergraph of graphs $\mathbb{G}_0, \mathbb{G}_1, \ldots, \mathbb{G}_n$ from Eq. (22) denoted by $\mathbb{G}_0 \vee \mathbb{G}_1 \vee \mathbb{G}_2 \vee \cdots \vee \mathbb{G}_n$. Its construction is carried out successively going step-by-step from the left side to the right side in Eq. (22). We start with the construction of a common supergraph of $\mathbb{G}_0$ and $\mathbb{G}_1$, denoted by $\mathbb{G}_0 \vee \mathbb{G}_1$, in such a way that we add the just created edge/loop into the graph $\mathbb{G}_0$. Here we have to distinguish between two different cases. First, the created edge/loop is already present in $\mathbb{G}_0$, then we only need to increase the multiplicity of this edge/loop by 1. Second, the related edge/loop does not belong to the edge set of $\mathbb{G}_0$, then its edge set is enlarged by this edge/loop and its multiplicity is put equal to 1. In a similar way we construct in the forthcoming step a common supergraph of the just constructed $\mathbb{G}_0 \vee \mathbb{G}_1$ and $\mathbb{G}_2$, the resulting common supergraph will be denoted by $\mathbb{G}_0 \vee \mathbb{G}_1 \vee \mathbb{G}_2$. Finally, the constructed common supergraph $\mathbb{G}_0 \vee \mathbb{G}_1 \vee \mathbb{G}_2 \vee \cdots \vee \mathbb{G}_n$ is composed of those edges/loops that were initially present in $\mathbb{G}_0$ and also all created edges/loops going successively (in Eq. (22)) from $\mathbb{G}_0$ to $\mathbb{G}_n$ through intermediates $\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_{n-1}$. This means that the cardinality of this common supergraph is equal to the cardinality of edge set $E(\mathbb{G}_0)$ (where $|E(\mathbb{G}_0)| = q$) plus the length of minimum path of Eq. (22), that is

$|\mathbb{G}_0 \vee \mathbb{G}_1 \vee \mathbb{G}_2 \vee \cdots \vee \mathbb{G}_n| = q + n$. Since integer n is in fact equal to the reaction distance between graph $\mathbb{G}$ and $\mathbb{G}'$, we arrive at the following theorem.

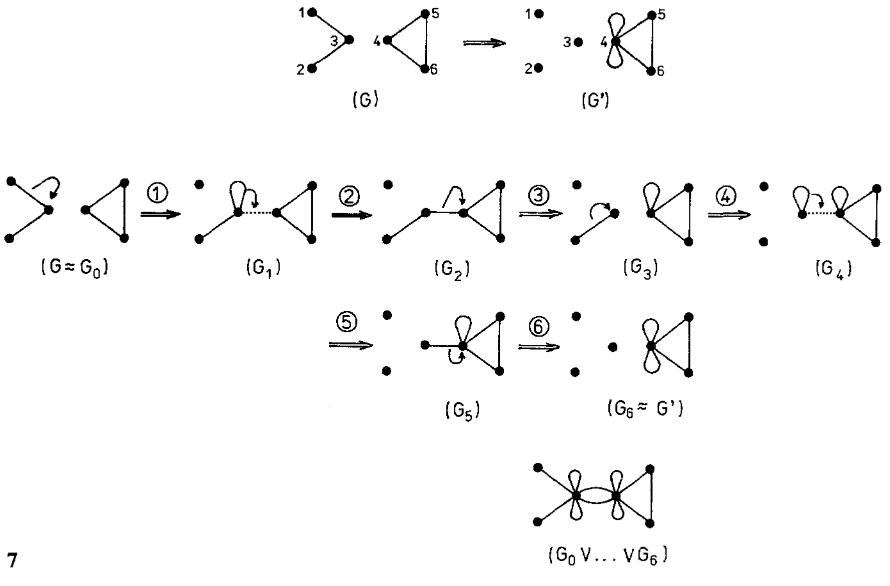
Theorem 5. The reaction distance between a pair of graphs $\mathbb{G}$, $\mathbb{G}' \in \mathscr{F}_{pq}$ is:

$$RD(\mathbb{G}, \mathbb{G}') = |\mathbb{G}_0 \vee \mathbb{G}_1 \vee \mathbb{G}_2 \vee \cdots \vee \mathbb{G}_n| - q, \tag{24}$$

where $\mathbb{G}_0 \approx \mathbb{G}, \mathbb{G}_1, \ldots, \mathbb{G}_{n-1}, \mathbb{G}_n \approx \mathbb{G}'$ is a sequence of molecular graphs from a shortest path connecting $\mathbb{G}$ and $\mathbb{G}'$ in $\mathbb{G}_{pq}(\mathscr{R})$. □

This theorem represents the principal result of this section. It demonstrates that the reaction distance, initially determined as a graph distance over the corresponding graph of distances $\mathbb{G}_{pq}(\mathscr{R})$, may be alternatively determined by Eq. (24) with the help of the concept of a common supergraph. Unfortunately, this common supergraph is determined not only by the molecular graphs for which the reaction distance is evaluated but also by graphs – intermediates which lie on a shortest path connecting the graphs $\mathbb{G}$ and $\mathbb{G}'$ in $\mathbb{G}_{pq}(\mathscr{R})$.

Example 4. Let us consider a pair of graphs $\mathbb{G}$ and $\mathbb{G}'$ belonging to the family $\mathscr{F}_{6,5}$, see Fig. 7. Their transformation $\mathbb{G} \Rightarrow \mathbb{G}'$ may be decomposed in a minimum sequence of six elementary tranformation determined by the reaction graphs from Eq. (21), i.e. their reaction distance is $RD(\mathbb{G}, \mathbb{G}') = 6$. The corresponding common supergraphs of all seven graphs (one educt graph $\mathbb{G} \approx \mathbb{G}_0$, five intermediate graphs $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_3, \mathbb{G}_4, \mathbb{G}_5$, and one product graph $\mathbb{G}_6 \approx \mathbb{G}'$) are displayed in the fourth row in Fig. 7. The cardinality of the common supergraph $\mathbb{G}_0 \vee \cdots \vee \mathbb{G}_6$ is equal to 11. According to Eq. (25) the reaction distance is $RD(\mathbb{G}, \mathbb{G}') = 11 - 5 = 6$. □



7

Fig. 7. Decomposition of the transformation $\mathbb{G} \Rightarrow \mathbb{G}'$ onto a minimum sequence of six prototype transformations assigned to reaction graphs from the set $\mathscr{R}$ specified by Eq. (25). The common supergraph $\mathbb{G}_0 \vee \cdots \vee \mathbb{G}_6$ is displayed in the fourth row

Chemical distance. Chemical distance was independently determined in the framework of the maximum common subgraph by Johnson [19] and Baláž et al. [20–22]. The set $\mathscr{R}$ is composed of all prototype reaction graphs listed in Fig. 3:

$$\mathscr{R} = [\mathbb{G}_R^{(i)};\ 1 \leqslant i \leqslant 7\} \tag{25}$$

Since the set determined by Eq. (21) is a subset of this set, then, according to Theorems 3 and 4, the graph of distances $\mathbb{G}_{pq}(\mathscr{R})$ assigned to Eq. (25) is connected. The distance $D(\mathbb{G}, \mathbb{G}')$ determined over the graph $\mathbb{G}_{pq}(\mathscr{R})$ will be called the chemical distance and denoted by $CD(\mathbb{G}, \mathbb{G}')$. Following Theorem 3, chemical distance should be bounded from above by the reaction distance:

$$0 \leqslant CD(\mathbb{G}, \mathbb{G}') \leqslant RD(\mathbb{G}, \mathbb{G}') \tag{26}$$

that is, the reaction distance represents an exact upper bound of the chemical distance. From the above inequality immediately follows that if $RD = 1$, then also $CD = 1$, formally:

$$RD(\mathbb{G}, \mathbb{G}') = 1 \Rightarrow CD(\mathbb{G}, \mathbb{G}') = 1 \tag{27}$$

Unfortunately, the implication of Eq. (27) could not be reversed; assuming that $CD(\mathbb{G}, \mathbb{G}') = 1$, then [22] $1 \leqslant RD(\mathbb{G}, \mathbb{G}') \leqslant 4$. This property may be straightforwardly generalized for a pair of arbitrary molecular graphs from $\mathscr{F}_{pq}$, and we get $RD(\mathbb{G}, \mathbb{G}') \leqslant 4CD(\mathbb{G}, \mathbb{G}')$. Introducing this inequality into Eq. (26) we arrive at the very important relation between reaction and chemical distance [18]:

$$CD(\mathbb{G}, \mathbb{G}') \leqslant RD(\mathbb{G}, \mathbb{G}') \leqslant 4CD(\mathbb{G}, \mathbb{G}') \tag{28}$$

Example 5. The graph of distances $\mathbb{G}_{2,3}(\mathscr{R})$ determined with respect to the set of Eq. (25) is displayed in Fig. 8. We see that the graph of distances studied in Example 3 (see Fig. 6) is a subgraph of the present graph of distances (cf. Theorem 3). This property immediately implies the inequality of Eq. (26). The matrix of chemical distances between graphs from $\mathscr{F}_{2,3}$ is:

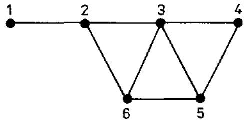
$$CD = \begin{pmatrix} 0 & 1 & 2 & 3 & 3 & 2 \\ 1 & 0 & 1 & 2 & 2 & 1 \\ 2 & 1 & 0 & 1 & 1 & 1 \\ 3 & 2 & 1 & 0 & 1 & 2 \\ 3 & 2 & 1 & 1 & 0 & 1 \\ 2 & 1 & 1 & 2 & 1 & 0 \end{pmatrix}. \qquad \square$$

Example 6. On the first sight it may seem that the right side inequality of Eq. (28) considerably overestimates the upper bound of the reaction distance. Let us study two isomeric graphs $\mathbb{G}_1$ and $\mathbb{G}_2$ in Fig. 9. Their distances are $RD(\mathbb{G}_1, \mathbb{G}_2) = 4$ and $CD(\mathbb{G}_1, \mathbb{G}_2) = 1$, that is the term $4CD(\mathbb{G}_1, \mathbb{G}_2)$ represents the exact upper bound of $RD(\mathbb{G}_1, \mathbb{G}_2)$. □

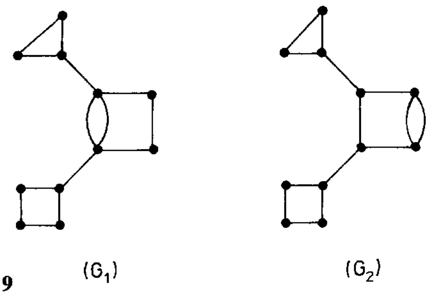
Theorem 6. The chemical distance $CD(\mathbb{G}, \mathbb{G}')$ between graphs from $\mathscr{F}_{pq}$ is determined by the following two equivalent ways, namely:

$$CD(\mathbb{G}, \mathbb{G}') = \begin{cases} q - |\mathbb{G} \cap \mathbb{G}'| \\ |\mathbb{G} \cup \mathbb{G}'| - q. \end{cases} \qquad \square \tag{29}$$

The proof of this theorem may be done in a similar way as the proof of Theorem 5. Let us consider a pair of graphs $\mathbb{G}, \mathbb{G}' \in \mathscr{F}_{pq}$ and let a shortest path in $\mathbb{G}_{pq}(\mathscr{R})$ connecting these vertices be equal to n, i.e. $CD(\mathbb{G}, \mathbb{G}') = n > 0$. It



8



9

Fig. 8. Graph of distances determined over the family $\mathscr{F}_{2,3}$ and with respect to the set $\mathscr{R}$ of prototype reaction graphs of Eq. (25). The length of the shortest path between vertices – molecular graphs $\mathbb{G}$, $\mathbb{G}' \in \mathscr{F}_{2,3}$ is equal to the chemical distance $CD(\mathbb{G}, \mathbb{G}')$ between them

Fig. 9. Two isomeric molecular graphs with chemical distance $CD(\mathbb{G}_1, \mathbb{G}_2) = 1$ and reaction distance $RD(\mathbb{G}_1, \mathbb{G}_2) = 4$

means that there exists a sequence (Eq. (22)) of isomeric molecular graphs composed of $(n-1)$ intermediate graphs $\mathbb{G}_1, \mathbb{G}_2, \ldots, \mathbb{G}_{n-1}$. Similarly as for the reaction distance, these graphs are mutually recurrently related by transformation Eq. (23). The corresponding common supergraph of $\mathbb{G}_0, \mathbb{G}_1, \ldots, \mathbb{G}_{n-1}, \mathbb{G}_n$ is constructed by making use of the same procedure as that used in the previous subsection for reaction distance [see the comment below Eq. (23)]:

$$CD(\mathbb{G}, \mathbb{G}') = |\mathbb{G}_0 \vee \mathbb{G}_1 \vee \mathbb{G}_2 \vee \cdots \vee \mathbb{G}_n| - q \tag{30}$$

Since the set $\mathscr{R}$ is composed of all reaction graphs in Fig. 3, each elementary transformation corresponds to a process of creation of an edge/loop which is also appearing in the product graph $\mathbb{G}'$. This means that the common supergraph $\mathbb{G}_0 \vee \mathbb{G}_1 \vee \mathbb{G}_2 \vee \ldots \vee \mathbb{G}_n$ should be isomorphic to a minimum common supergraph $\mathbb{G} \cup \mathbb{G}'$, and consequently:

$$CD(\mathbb{G}, \mathbb{G}') = |\mathbb{G} \cup \mathbb{G}'| - q \tag{31}$$

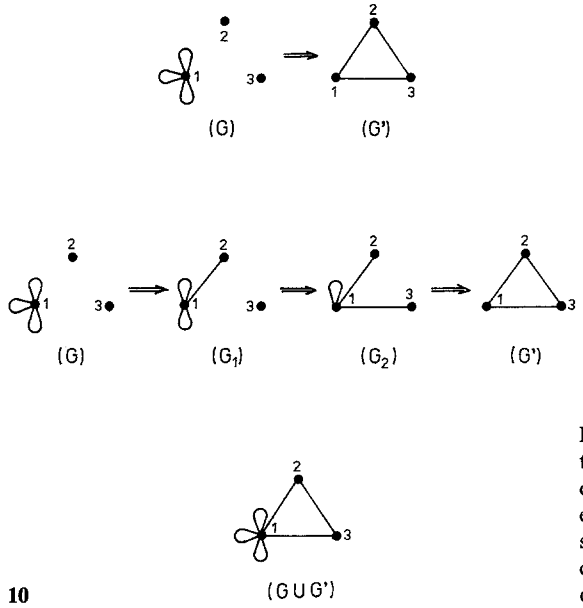
Applying simple set-theoretical considerations we get that the maximum common subgraph and the minimum common supergraph are related by:

$$|\mathbb{G} \cup \mathbb{G}'| + |\mathbb{G} \cap \mathbb{G}'| = |\mathbb{G}| + |\mathbb{G}'| = 2q \tag{32}$$

Introducing Eq. (32) into Eq. (31) we arrive at the first row of Eq. (29) whereas Eq. (31) is equal to the second row of Eq. (29).

Theorem 6 determines the chemical distance between pairs of isomeric molecular graphs by making use of the concept of their maximum (minimum) common subgraphs (supergraphs). It means that this distance depends explicitly only on the graphs for which the distance is evaluated and, in constrast to the reaction distance, it does not depend on intermediate graphs from the minimum path in the graph $\mathbb{G}_{pq}(\mathscr{R})$. Therefore, Eq. (29) may be used for an evaluation of chemical distance without the necessity to know the graph of distances $\mathbb{G}_{pq}(\mathscr{R})$ [25, 26].

Example 7. In order to illustrate the above proof of Theorem 6 we shall study the pair of graphs $\mathbb{G}, \mathbb{G}' \in \mathscr{F}_{pq}$, see Fig. 10. The transformation $\mathbb{G} \Rightarrow \mathbb{G}'$ may be decomposed into a minimum sequence of three elementary transformations, all of which are determined by reaction graphs from the set of Eq. (25), see Fig. 10.



10

Fig. 10. Illustrative example of transformation of molecular graphs decomposed into a sequence of elementary transformations from the set of Eq. (25). The minimum common supergraph of $\mathbb{G}$ and $\mathbb{G}'$ is displayed in the third row

We see that intermediate graphs $\mathbb{G}_1$ and $\mathbb{G}_2$ contain newly created edges which appear in the product graph $\mathbb{G}'$. The common supergraph displayed in the third row in Fig. 10 is isomorphic to the minimum common supergraph $\mathbb{G} \cup \mathbb{G}'$. Then, according to Eq. (29), the chemical distance between $\mathbb{G}$ and $\mathbb{G}'$ is determined by $CD(\mathbb{G}, \mathbb{G}') = 6 - 3 = 3$. □

4. Discussion

The reaction and chemical distances between molecular graphs taken from the same family of isomeric graphs are entities well suited to "measure" similarity or dissimilarity between them. These distances may be introduced in two quite different manners. In the first, the distances are determined as the length of minimum path which connects the graphs in the graph of distances. In the second, the distances are determined by resorting to the concept of common supergraph. Both these approaches are equivalent and offer simple, chemically relevant interpretation of distances. For instance, the reaction distance determines the minimum number of elementary reaction acts – associations and dissociations – that are necessary for transformation of an educt molecular graph onto a product molecular graph. It reflects Ingold's idea of a push-pull electron flow through a sequence of created and/or annihilated bonds. This combined with the famous principle of minimum structural change has successfully been used in organic chemistry for more than 100 years [9].

Chemical distance has not such a simple mechanistic interpretation. It reflects the principle of minimum structural change without referring to how the given transformation $\mathbb{G} \Rightarrow \mathbb{G}'$ is mechanistically realized. Its increasing value indicates that molecular graphs $\mathbb{G}$ and $\mathbb{G}'$ are more disimilar. In such a case more prototype reaction graphs must be applied to carry out the transformation $\mathbb{G} \Rightarrow \mathbb{G}'$.

As a byproduct of our graph-theoretical considerations new possibilities appear for the determination of a distance between molecular graphs. This is based on a choice of the set $\mathscr{R}$, its determination according to Eqs. (21) and (25) induces the reaction and chemical distances, respectively. Of course, there exists other interesting possibilities to determine the set $\mathscr{R}$ such that the corresponding graph of distances $\mathbb{G}_{pq}(\mathscr{R})$ is connected.

For instance, the set $\mathscr{R}$ determined by Eq. (21) may be enlarged by further reaction graphs in Fig. 3. The induced distance over the graph $\mathbb{G}_{pq}(\mathscr{R})$ may be interpreted as an intermediate metric bounded by reaction and chemical distances. Similarly, if the set $\mathscr{R}$ is composed of the prototype reaction graphs 3 and 4 shown in Fig. 3, then, according to Theorem 1, the graph $\mathbb{G}_{pq}(\mathscr{R})$ is connected and it induces a distance between molecular graphs. Unfortunately, these distances are of more graph-theoretical than chemical meaning, they represent an interesting extension of the original chemical idea to the problem of graph distances.

References

1. Dugundji J, Ugi I (1973) Top Curr Chem 39:19
2. Hendrickson JB (1976) Top Curr Chem 62:49
3. Zefirov NS, Tratch SS (1975) Zh Org Khim 11:25 (in Russian)
4. Zefirov NS (1987) Acc Chem Res 20:237 (and references therein)
5. Koča J, Kratochvíl M, Kvasnička V, Matyska L, Pospíchal J (1989) Synthon model of organic chemistry and synthesis design. Springer-Verlag, Berlin, Heidelberg, New York (and references therein)
6. Ugi I, Bauer J, Brandt J, Friedrich J, Gasteiger J, Jochum C, Schubert W (1979) Angew Chem Int Ed Engl 18:111
7. Hückel W (1952) Theoretische Grundlagen der organischen Chemie, Band 1. Akademie Verlag, Leipzig
8. Wheland GW (1962) Advanced organic chemistry. Wiley, New York
9. Kolbe M (1850) Liebigs Ann Chem 75:211; (1850) 76:1
10. Harary F (1969) Graph theory. Addison-Wesley, Reading, Mass
11. Balaban AT (ed) (1976) Chemical applications of graph theory. Academic Press, London
12. Fontain E, Bauer J, Ugi I (1987) Z Naturforsch 42b:889
13. Bauer J, Fontain E, Ugi I (1988) Anal Chim Acta 210:123
14. Ugi I, Bauer J, Baumgartner R, Fontain E, Forstmayer D, Lohberg S (1988) Pure Appl Chem 60:1573
15. Wochner M, Brandt J, von Scholey A, Ugi I (1988) Chimica 42:217
16. Pospíchal J, Kvasnička V (1990) Theor Chim Acta 76:423
17. Kvasnička V, Pospíchal J (1990) J Math Chem 5:309
18. Kvasnička V, Pospíchal J (1990) J Mol Struct (THEOCHEM) (in press)
19. Johnson MA (1985) In: Alavi Y, Chartrand G, Lesniak L, Wall C (eds) Graph theory and its applications to algorithms and computer science. Wiley, New York
20. Baláž V, Koča J, Kvasnička V, Sekanina M (1986) Časopis pro pěstováni matematiky 111:431
21. Baláž V, Kvasnička V, Pospíchal J (1989) Časopis pro pěstováni matematiky 114:155
22. Baláž V, Kvasnička V, Pospíchal J (1990) Discr Appl Math (in press)
23. Koča J (1988) Coll Czech Chem Comm 53:3119
24. Ingold CK (1969) Structure and mechanisms in organic chemistry. Cornell Univ Press, Ithaca
25. Nicholson V, Tsai CC, Johnson M, Naim M (1987) In: King RB, Rouvray DH (eds) Graph theory and topology in chemistry. Elsevier, Amsterdam
26. Kvasnička V, Pospíchal J (1990) Reports Mol Theor (in press)
27. Harary F (1953) Michigan Math J 2:142