Two metrics in a graph theory modeling of organic chemistry*

Vladimír Baláž, Vladimír Kvasnička and Jiří Pospíchal

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

Received 24 January 1989
Revised 17 November 1989

Abstract


Two different metrics in the graph theory model of organic chemistry are defined. The chemical distance between two graphs (molecules) taken from the same family of isomeric graphs $S_{pq}$ (where $p$ is a number of vertices and $q$ is a number of edges and loops) is based on the maximal common subgraph. The reaction distance assigned to a transformation $G = G'$, where $G, G' \in S_{pq}$, is equal to the minimal number of the so-called elementary transformations that are necessary for the transformation of $G$ onto $G'$. Both the suggested metrics reflect formally many known fundamental chemical rules, in particular, the principle of minimal structural change.

Keywords. Chemical distance, reaction distance, graph theory.

1. Introduction

The concepts of graph and molecular graph are widely used in the mathematical chemistry [1-3,10,15,22,23]. The graph theory provides very effective formal tools how to formalize the organic chemistry. Up to now, most works dealing with the applications of graph theory to organic chemistry are turned mainly towards the enumeration problem, i.e., they enumerate and/or generate the graphs with simple correspondence to structural types of molecules. Another very fruitful field of the application of graph theory concepts and notions to organic chemistry is to quantify a similarity (or dissimilarity) between two graphs by making use of different metrics [7,8,12-14,21,27-29]. Recently, we have suggested [4,5] the so-called edge distance

* Dedicated to Professor Milan Kratochvíl in honor of his 65th birthday.

0166-218X/92/$05.00 © 1992 — Elsevier Science Publishers B.V. All rights reserved
(called in this work the chemical distance) (cf. also [12,13]). This metric reflects in a formal way the somewhat vaguely formulated principle of minimal structural change. It states that a minimal number of chemical bonds are generally cancelled/form ed during a chemical reaction [11,25]. In the graph theory approach it means that a transformation of a graph G onto another graph G' (G = G') is carried out in such a way that a greatest common part of G and G' (expressed by their maximal common subgraph) is saved and only a minimal number of edges/loops are cancelled/form ed in order to get G' from G. The second type of metric, called the reaction metric, requires that a transformation G = G' is carried through a sequence of minimal number of the so-called elementary transformations. Under the term “elementary transformation” we understand such a transformation in which one edge (loop) is cancelled and one loop (edge) is formed, and moreover, the edge and the loop are adjacent (i.e., they have a vertex in common). Similarly, the concept of reaction metric is very suitable for the graph theory formulation of many different aspects of chemical reactions. Both metrics initially suggested in the framework of the graph theory modeling of organic chemistry, might be of more general applicability than merely for purposes of organic chemistry. They offer new and nontraditional look at relatively old and still unsolved problems of graph theory, e.g. the graph reconstruction problem [6].

2. Basic concepts

A **vertex set** \( V = \{v_1, v_2, \ldots, v_N\} \) is a nonempty set composed of \( N \) vertices \( v_1, v_2, \ldots, v_N \). 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. An **edge set** \( E = \{e_1, e_2, \ldots, e_M\} \), where \( e_i = [v_j, v_k] \), associated with the vertex set \( V \), is a set of \( M \) edges \( e_1, e_2, \ldots, e_M \) 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 multiplicity \( u \) is a set of \( u \) loops incident with the same vertex. A **loop set** \( L = \{l_1, l_2, \ldots, l_P\} \), 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 \( v_i \).

**Definition 2.1.** A graph is an ordered triple

\[
G = (V, E, L),
\]  

(2.1)

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).

From the graph $G=(V,E,L)$ we may form its subgraph, which is also the graph, by successive applications of one of the following two alternative 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 2.2.** 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, 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 2.3.** 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 a 1-1 mapping $\psi : V_1 \to V_2$ which induces two 1-1 mappings $\chi : E_1 \to E_2$ and $\chi' : L_1 \to L_2$ that conserve the incidences of edges and loops, respectively.

One of the most fundamental notions of organic chemistry is that of isomerism. Two or more molecules are isomeric iff they contain the same number of atoms and the same number of valence electrons but differ in the structural arrangement of atoms.

**Definition 2.4.** Two graphs $G_1=(V_1,E_1,L_1)$ and $G_2=(V_2,E_2,L_2)$ are isomeric $(G_1 \cong G_2)$ iff

$$|V_1|=|V_2|,$$

$$|E_1|+|L_1|=|E_2|+|L_2|,$$ (2.2a) (2.2b)

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

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

1. Reflexive law, $G=G$.
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 isomorphic graphs.
Definition 2.5. The family of isomeric graphs $\mathcal{F}_{pq}$ is composed of all possible non-isomorphic graphs with $p$ vertices and $q$ edges and loops,

$$\mathcal{F}_{pq} = \{G = (V,E,L); |V| = p \text{ and } |E| + |L| = q\}. \quad (2.3)$$

Example 2.6. The family $\mathcal{F}_{2,3}$ of isomeric graphs composed of two vertices and three edges and/or loops contains the following six graphs:

3. Chemical distance

For the graph theory model of organic chemistry the notion of chemical distance was introduced by the present authors [4,5]. Its definition is closely related to the Dugundji and Ugi [9] matrix model of organic chemistry, where the chemical distance is determined as the Hamming norm of the difference of adjacency matrices.

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 subgraphs $G_1 \subseteq G_1$ and $G_2 \subseteq G_2$, that is $G \cong G_1 \approx G_2$. 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 $G_2$. 
Definition 3.2. The chemical distance 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 \( G_1 \cap G_2 = (V_{12}, E_{12}, L_{12}) \) is determined by

\[
CD(G_1, G_2) = |E_1| + |E_2| + |L_1| + |L_2| - 2|E_{12}| - 2|L_{12}|.
\] (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 the maximal common subgraph. It enables to “measure” a similarity (or dissimilarity) of two different isomeric graphs, its increasing value indicates that the graphs are more dissimilar.

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 \( |L_{12}| = 1 \), then \( CD(G_1, G_2) = 7 + 7 - 2 \cdot 5 - 2 \cdot 1 = 2 \).

Theorem 3.4. The chemical distance \( CD(G_1, G_2) \) for isomeric graphs from the family \( \mathcal{F}_{pq} \) is a metric, that is the following three properties are satisfied:

1. Positive semidefiniteness,
   \[
   CD(G_1, G_2) \geq 0 \quad (= 0 \text{ only for } G_1 \approx G_2). \tag{3.2a}
   \]

2. Symmetry,
   \[
   CD(G_1, G_2) = CD(G_2, G_1). \tag{3.2b}
   \]

3. Triangle inequality,
   \[
   CD(G_1, G_2) + CD(G_2, G_3) \geq CD(G_1, G_3). \tag{3.2c}
   \]
The first two properties of the chemical distance follow immediately from Definition 3.2. The triangle inequality (3.2c) was proved by two independent methods by the present authors [4,5].

We have to emphasize that the chemical distance was defined via the notion of maximal common subgraph. Hence, in order to evaluate the chemical distance \( \text{CD}(G_1, G_2) \) we have to know their maximal common subgraph \( G_1 \cap G_2 \). McGregor [20] has suggested a back-track searching algorithm for the construction of the maximal common subgraph of two graphs. His method involves, at worst, \( N! \) backtrack searches. For molecular graphs, where the vertices are evaluated by the chemical symbols, this approach could be substantially accelerated by making use of some effective heuristics (cf. [26]).

**Theorem 3.5.** The chemical distance \( \text{CD}(G_1, G_2) \) for \( G_1, G_2 \in \mathcal{F}_{pq} \) is even and bounded from above,

\[
\text{CD}(G_1, G_2) = 2k \leq 2q, \tag{3.3}
\]

where \( k = 0, 1, 2, \ldots, q \).

Since \( G_i = (V_i, E_i, L_i) \), \( G_2 = (V_2, E_2, L_2) \in \mathcal{F}_{pq} \) we have

\[
|E_1| + |L_1| = |E_2| + |L_2| = q.
\]

Introducing this relation in (3.1) we get

\[
\text{CD}(G_1, G_2) = 2(q - |E_{12}| - |L_{12}|), \tag{3.4}
\]

which 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 \( \mathcal{G}^\text{CD}_{pq} \). Its vertex set is identified with the family \( \mathcal{F}_{pq} \), i.e., each vertex of \( \mathcal{G}^\text{CD}_{pq} \) 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 3.6.** The matrix of chemical distances between graphs from the family \( \mathcal{F}_{2,3} \) (see Example 3.1) is a symmetric matrix with diagonal entries equal to zero,

\[
\begin{pmatrix}
0 & 2 & 4 & 6 & 6 & 4 \\
2 & 0 & 2 & 4 & 4 & 2 \\
4 & 2 & 0 & 2 & 2 & 2 \\
6 & 4 & 2 & 0 & 2 & 4 \\
6 & 4 & 2 & 2 & 0 & 4 \\
4 & 2 & 2 & 4 & 4 & 0
\end{pmatrix}
\]

The graph of chemical distances constructed over the family \( \mathcal{F}_{2,3} \) contains six vertices and seven edges,
Theorem 3.7. The graph $\mathcal{G}_{pq}^{\text{CD}}$ is connected.

Theorem 3.8. The chemical distance between graphs from the family $\mathcal{F}_{pq}$ is double the graph distance between them in the graph $\mathcal{G}_{pq}^{\text{CD}}$.

Theorem 3.9. The necessary and sufficient condition that for a triple of distinct graphs $G_1, G_2, G_3 \in \mathcal{F}_{pq}$ the following relation is satisfied

\[ \text{CD}(G_1, G_2) + \text{CD}(G_2, G_3) = \text{CD}(G_1, G_3), \]  

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

In order to prove Theorems 3.7 and 3.8, let us consider a pair of graphs $G_1, G_2 \in \mathcal{F}_{pq}$ with chemical distance $\text{CD}(G_1, G_2) = 2n$ (we remember that according to Theorem 3.5 the chemical distance is a nonnegative even integer bounded from above). For $n = 1$ the graphs $G_1$ and $G_2$ are automatically adjacent by an edge in $\mathcal{G}_{pq}^{\text{CD}}$, hence their chemical distance is double the graph distance in $\mathcal{G}_{pq}^{\text{CD}}$. For $n > 1$ we can find a graph $G_3 = (V_3, E_3, L_3)$ satisfying $\text{CD}(G_1, G_3) = 2$ and $\text{CD}(G_3, G_2) = 2(n - 1)$. The graph $G_3$ is constructed in the following way: $V_3 = V_1$ and for construction of edge/loop sets we have four alternative possibilities,

1. $E_3 = (E_1 \setminus \{e'\}) \cup \{e''\}$, $L_3 = L_1$,
2. $E_3 = E_1 \setminus \{e'\}$, $L_3 = L_1 \cup \{l''\}$,
3. $E_3 = E_1 \cup \{e''\}$, $L_3 = L_1 \setminus \{l'\}$,
4. $E_3 = E_1$, $L_3 = (L_1 \setminus \{l'\}) \cup \{l''\}$.

Here, $e' \in E(G_1) \setminus E(G'_1)$, $e'' \in E(G_1)$, $G'_1 = (V_1, E(G_1) \cap G_2) \cup \{e''\}$, $G''_1 = G_1 \setminus G_2$, or $l' \in L(G_1) \setminus L(G'_1)$, $l'' \in L(G_1)$, $G''_1 = (V_1, E_1, L(G_1) \cap G_2) \cup \{l''\}$, $G''_1 = G_1 \setminus G_2$, where $G_1 \subseteq G_1$ and $G_1 = G_1 \cap G_2$. There must exist at least one of the above four constructions of $G_3$. If $2(n - 1) = 2$, then graphs $G_1$ and $G_2$ are connected by a path composed of two edges in $\mathcal{G}_{pq}^{\text{CD}}$. For $2(n - 1) > 2$, we take graphs $G_1$ and $G_3$ and construct another graph $G_x$ with $\text{CD}(G_3, G_x) = 2$ and $\text{CD}(G_x, G_2) = 2(n - 2)$. This process is repeated until a path between $G_1$ and $G_2$ is formed. Such a path may be constructed for all pairs of graphs in $\mathcal{F}_{pq}$, that is the graph $\mathcal{G}_{pq}^{\text{CD}}$ is connected.

Since the chemical distance between each pair of adjacent graphs in $\mathcal{G}_{pq}^{\text{CD}}$ is equal to 2 and the above constructed sequence of graphs gives a shortest path in $\mathcal{G}_{pq}^{\text{CD}}$ between $G_1$ and $G_2$, it is obvious that the chemical distance is double the graph distance in $\mathcal{G}_{pq}^{\text{CD}}$. 
As concerns Theorem 3.9, according to the above considerations, for each pair of graphs \( G_1 \) and \( G_2 \) there must exist a graph \( G_2 \) such that the chemical distance between \( G_1 \) and \( G_3 \) is exactly equal to the sum of chemical distances between graphs \( G_1 \) and \( G_2 \) and graphs \( G_2 \) and \( G_3 \). Moreover, the "intermediate" graph \( G_2 \) lies on the shortest path connecting the graphs \( G_1 \) and \( G_3 \) in \( \mathcal{G}_{pq}^{\text{CD}} \), which was to be proved. We have to emphasize that this theorem gives conditions under which equality holds in the triangle inequality (3.2c) in Theorem 3.4.

4. Reaction graph

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 transformation,

\[
G_1 \Rightarrow G_2,
\]

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 \( G_2 \), then \( G_1 \) and \( G_2 \) may be expressed as a "union" of two edge/loop disjoint subgraphs,

\[
G_1 = G'_1 \cup G''_1, \quad G_2 = G'_2 \cup G''_2,
\]

where the subgraphs \( G''_1 \subseteq G_1 \) and \( G''_2 \subseteq G_2 \) are isomorphic to \( G_1 \cap G_2 \), i.e., \( G''_1 = G''_2 = G_1 \cap G_2 \). We say that the transformation (4.1) changes the subgraph \( G'_1 \) onto the subgraph \( G'_2 \) whereas the subgraphs \( G''_1 \) and \( G''_2 \) remain intact. Hence, the transformation (4.1) can be simplified as follows

\[
G'_1 \Rightarrow G'_2,
\]

where we have omitted intact subgraphs of \( G_1 \) and \( G_2 \). Let us consider a graph \( G'_2 = (V'_2, E'_2, L'_2) \), which is isomorphic to \( G'_2 \), i.e., \( G'_2 = G'_2 \), and it is constructed over the same vertex set as the graph \( G'_1 = (V'_1, E'_1, L'_1) \).

**Definition 4.1.** The reaction graph [19] \( G_R \) assigned to the transformation (4.1) is an ordered 4-tuple

\[
G_R = (V_R = V'_1, E_R, L_R, \psi),
\]

\[
E_R = E'_1 \cup E'_2, \quad L_R = L'_1 \cup L'_2.
\]

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

\[
\psi(x) = \begin{cases} 
-1 & \text{for } x \in E'_1 \cup L'_1, \\
+1 & \text{for } x \in E'_2 \cup L'_2.
\end{cases}
\]

The notion of reaction graph (assigned to a fixed transformation \( G_1 \Rightarrow G_2 \)) enables us to "algebraicize" the transformation as follows:
\[ G_1 + G_R = \bar{G}_2 \cong G_2. \] (4.5)

The binary operation "\( + \)" is interpreted in the framework of set theory formalism as the symmetric difference: \( A + B = (A \setminus B) \cup (B \setminus A) \). Applying this operation the edge and loop sets of \( \bar{G}_2 \) are:

\[ \bar{E}_2 = E_1 + E_R, \quad \bar{L}_2 = L_1 + L_R. \] (4.6)

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

**Example 4.2.** Consider the transformation \( G_1 \Rightarrow G_6 \), where \( G_1 \) and \( G_6 \) are graphs specified in Example 3.1, diagrammatically

![Diagram](image)

Their maximal common subgraph \( G_1 \cap G_6 \) is

![Diagram](image)

This graph represents the so-called intact part of \( G_1 \) and \( G_6 \) remaining invariant during the transformation \( G_1 \Rightarrow G_6 \). The simplified transformation (4.3) looks like

![Diagram](image)

Unifying both sides of this transformation at one graph and evaluating its edges/loops from the left (right) side by \(-1\) \((+1)\), we arrive at the reaction graph \( G_R \) assigned to the transformation \( G_1 \Rightarrow G_6 \),

![Diagram](image)

Its interpretation is very simple, during the transformation \( G_1 \Rightarrow G_6 \) two edges are cancelled and two loops are formed.

**Theorem 4.3.** The chemical distance between two graphs \( G, G' \in \mathcal{F}_{pq} \) satisfies

\[ \text{CD}(G, G') = |E_R| + |L_R|, \] (4.7)
where $E_R$ ($L_R$) is an edge (loop) set of the reaction graph $G_R$ assigned to the transformation $G \Rightarrow G'$.

According to the determination of the reaction graph (see Definition 4.1), the cardinalities of its edge and loop sets are determined by

\begin{align}
|E_R| &= |E(G)| + |E(G')| - 2|E(G \cap G')|, \\
|L_R| &= |L(G)| + |L(G')| - 2|L(G \cap G')|,
\end{align}

(4.8a, 4.8b)

\begin{equation}
|E_R| + |L_R| = 2(q - |E(G \cap G')| - |L(G \cap G')|).
\end{equation}

Comparing this relation with (3.4) we immediately obtain the equality (4.7). The right-hand side of (4.7) is minimal, which follows from the construction of the reaction graph $G_R$, where the corresponding maximal common subgraph $G \cap G'$ is determined by the requirement that it contains a maximal number of edges and loops.

Theorem 4.3 has an interesting chemical interpretation. Following Ugi et al. [9,19,20,24,26], the reaction graph $G_R$ assigned to the transformation $G \Rightarrow G'$ is constructed in such a way that a minimal number of edges/loops is cancelled and formed to get the product graph $G'$ from the educt graph $G$. They called this property the principle of minimal chemical distance and it might be of value as an effective heuristic for looking for a 'mechanistic way' of the transformation $G \Rightarrow G'$. This transformation may be carried out by many accessible manners but only those that are satisfying the condition $|E_R| + |L_R| = \min$ are acceptable from the standpoint of organic chemistry.

5. Reaction distance

In the previous section we have determined the reaction graph $G_R$ for a transformation $G \Rightarrow G'$ on the basis of their maximal common subgraph $G \cap G'$. There exists another alternative possibility [16-18] how to construct this reaction graph, to require that the transformation $G \Rightarrow G'$ is carried out by a sequence composed of a minimal number of the so-called elementary transformations, the number of which determines a reaction distance between the graphs $G$ and $G'$, and this distance induces a new metric for a given family of isomeric graphs.

Let us study the graph $G = (V, E, L) \in \mathcal{F}_{pq}$; we introduce two types of elementary transformations of $G$ onto $G' = (V', E', L') \in \mathcal{F}_{pq}$.

1. The graph $G$ contains the edge $[v_i, v_j]$, the elementary transformation $\alpha_{ij}$ "dissociates" the edge $[v_i, v_j]$ onto a loop $[v_j, v_j]$, schematically

![Diagram](image-url)
A graph theory modeling of organic chemistry

The resulting graph $G'$, isomorphic to a graph from $\mathcal{G}_{pq}$, has edge and loop sets determined by

$$E' = E \setminus \{[u_i, v_j]\}, \quad L' = L \cup \{[v_j, v_i]\}. \quad (5.1b)$$

(2) The graph $G$ contains the loop $[u_i, v_i]$ and the vertex $v_j$, the elementary transformation $\beta_{ij}$ "associates" the loop $[v_i, v_i]$ onto an edge $[v_i, v_j]$, schematically

The resulting graph $G'$, isomorphic to a graph from $\mathcal{G}_{pq}$, has edge and loop sets determined by

$$E' = E \cup \{[v_i, v_j]\}, \quad L' = L \setminus \{[v_i, v_i]\}. \quad (5.2b)$$

The elementary transformation $\beta_{ij}$ is formally considered as a "retrotransformation" with respect to the transformation $\alpha_{ij}$. The above elementary transformations are simply visualized by the following two reaction graphs,

In a similar way as was done in Section 3 we construct for a fixed family $\mathcal{G}_{pq}$ of isomeric graphs the so-called graph of reaction distances $\mathcal{G}^{RD}_{pq}$. Its vertex set is again identified with the family $\mathcal{G}_{pq}$. Two distinct vertices (corresponding to nonisomorphic graphs $G, G' \in \mathcal{G}_{pq}$) are connected by an edge iff there exists an elementary transformation $\xi = \alpha, \beta$ such that $G \xrightarrow{\xi} G'$.

**Example 5.1.** Study the graphs from $\mathcal{G}_{2,3}$ (cf. Example 3.1). For the following pairs of graphs there exist elementary transformations such that a member of the given pair is transformed onto another one from the same pair:

$$\begin{align*}
\mathcal{G}_1 & \quad \xrightarrow{\alpha_{21}} \quad \mathcal{G}_2 \\
\mathcal{G}_2 & \quad \xrightarrow{\alpha_{21}} \quad \mathcal{G}_3
\end{align*}$$
This means that the graph $\mathcal{G}_{pq}^{\text{RD}}$ looks like this:

\[ \begin{array}{cccccc}
1 & 2 & 3 & 4 & 5 & 6 \\
2 & 1 & 6 & 5 & 4 & 3 \\
3 & 2 & 1 & 0 & 2 & 3 \\
4 & 3 & 2 & 1 & 0 & 3 \\
5 & 4 & 3 & 2 & 1 & 0 \\
6 & 5 & 4 & 3 & 2 & 1 \\
\end{array} \]

**Theorem 5.2.** The graph $\mathcal{G}_{pq}^{\text{RD}}$ is connected.

Since the graphs from the family $\mathcal{F}_{pq}$ are in general pseudographs, for an arbitrary pair $G, G' \in \mathcal{F}_{pq}$ there must exist a finite sequence of elementary transformations "modifying" the graph $G$ onto a graph isomorphic to the graph $G'$. Its exact proof may be done following the slight modification of the proof given by Chartrand et al. [7] for their edge rotational distance.

**Definition 5.3.** The reaction distance between two graphs $G, G' \in \mathcal{F}_{pq}$, denoted by $\text{RD}(G, G')$, is equal to the graph distance between them in $\mathcal{G}_{pq}^{\text{RD}}$.

**Example 5.4.** The matrix of graph distances in $\mathcal{G}_{2,3}^{\text{RD}}$ (see Example 5.1) is a symmetric matrix with vanishing diagonal entries:

\[
\begin{pmatrix}
0 & 1 & 2 & 3 & 3 & 2 \\
1 & 0 & 1 & 2 & 2 & 1 \\
2 & 1 & 0 & 1 & 1 & 2 \\
3 & 2 & 1 & 0 & 2 & 3 \\
3 & 2 & 1 & 2 & 0 & 3 \\
2 & 1 & 2 & 3 & 3 & 0 \\
\end{pmatrix}
\]

According to Definition 5.3 this matrix is equivalent to the matrix of reaction distances for the family $\mathcal{F}_{2,3}$. 

Theorem 5.5. The reaction distance \( RD(G_1, G_2) \) for isomeric graphs from the family \( \mathbb{F}_{pq} \) is a metric, the following three properties are satisfied:

1. **Positive semidefiniteness,**
   \[
   RD(G_1, G_2) \geq 0 \quad (= 0 \text{ only for } G_1 = G_2). 
   \]  

2. **Symmetry,**
   \[
   RD(G_1, G_2) = RD(G_2, G_1). 
   \]  

3. **Triangle inequality,**
   \[
   RD(G_1, G_3) + RD(G_3, G_2) \geq RD(G_1, G_2), 
   \]  

where the equality is satisfied iff the graph \( G_3 \) lies on a shortest path connecting the graphs \( G_1 \) and \( G_2 \).

The proof of this theorem is obvious, it immediately follows from the properties of the graph distance.

The chemical and reaction distances between graphs from the same family \( \mathbb{F}_{pq} \) are not simply related, the following two theorems are dealing with this problem.

Theorem 5.6. If for a pair of graphs \( G, G' \in \mathbb{F}_{pq} \) the reaction distance is equal to one, \( RD(G, G') = 1 \), then the chemical distance between them is equal to two, \( CD(G, G') = 2 \).

Let us assume that \( RD(G, G') = 1 \), then there exists an elementary transformation \( \xi = \alpha, \beta \) such that \( G \not\sim G' \). This elementary transformation corresponds to the reaction graphs (5.3a) and (5.3b), respectively. Applying Theorem 4.3, the chemical distance must be 2, which was to be proved.

The above Theorem 5.6 has the following two simple corollaries.

Corollary 5.7. The graph of reaction distances is a subgraph of the graph of chemical distances,

\[
\mathcal{G}_{pq}^{RD} \subset \mathcal{G}_{pq}^{CD}. 
\]  
(5.5)

According to Theorem 5.6 an edge from \( \mathcal{G}_{pq}^{RD} \) must be an edge of the graph \( \mathcal{G}_{pq}^{CD} \).

Corollary 5.8. The chemical and reaction distances for an arbitrary pair of graphs \( G, G' \in \mathbb{F}_{pq} \) are related by

\[
\frac{1}{2} CD(G, G') \leq RD(G, G'). 
\]  
(5.6)

Since the graph \( \mathcal{G}_{pq}^{RD} \) is a subgraph of the graph \( \mathcal{G}_{pq}^{CD} \) (see Corollary 5.7) a graph distance between vertices (corresponding to the graphs \( G \) and \( G' \)) in the graph \( \mathcal{G}_{pq}^{CD} \)
Theorem 5.9. If for a pair of graphs $G, G' \in \mathcal{F}_{pq}$ the chemical distance is equal to 2, $CD(G, G') = 2$, then the reaction distance between them is bounded by

$$1 \leq RD(G, G') \leq 4. \quad (5.7)$$

Since $CD(G, G') = 2$, the reaction graph $G_{R}$ assigned to transformation $G \Rightarrow G'$ is one of the respective five forms listed in Fig. 1. This list of "elementary" reaction graphs is complete, this may be verified by simple combinatorial considerations involving all possibilities which can appear between graphs from the same family $\mathcal{F}_{pq}$.

Case 1: The reaction graph is $A$ or $A'$; then the reaction distance is $RD(G, G') = 1$.

Case 2: The reaction graph is $B$; then the reaction distance is $RD(G, G') = 2$.

Case 3: The reaction graph is $C$ or $C'$; then the reaction distance is $RD(G, G') = 3$.

Case 4: The reaction graph is $D$; then the reaction distance is $RD(G, G') = 4$.

Case 5: The reaction graph is $E$; then the reaction distance is $RD(G, G') = 2$.

These all possibilities for $G_{R}$ imply that the reaction distance is bounded by (5.7).

We have to note that the reaction distance for simple graphs (without multiedges and loops) is simply related to the edge rotational distance introduced by Chartrand et al. [7]. In particular, the reaction distance is double the edge rotational distance.

Remark 5.10. We have defined two different distances between graphs from the same family $\mathcal{F}_{pq}$. Unfortunately, the chemical and reaction distances are not "isometric", therefore they may produce different reaction graphs. These distances
can be combined into the so-called intermediate distance,

\[ ID_\xi(G, G') = \lambda \text{RD}(G, G') + (1 - \lambda) \frac{1}{2} \text{CD}(G, G'), \quad (5.8) \]

where \(0 \leq \lambda \leq 1\) and the chemical distance CD is on the right side multiplied by a factor \(\frac{1}{2}\) (we remember that the chemical distance is an even number, see Theorem 3.5). For \(\lambda = 0\) (\(\lambda = 1\)) the intermediate distance is equal to the one half of the chemical distance (reaction distance), i.e., \(\text{ID}_0(G, G') = \frac{1}{2} \text{CD}(G, G')\) and \(\text{ID}_1(G, G') = \text{RD}(G, G')\). For \(0 < \lambda < 1\) the intermediate distance is equal to a convex combination of the one half of chemical distance and the reaction distance. According to Corollary 5.8 we get

\[ \frac{1}{2} \text{CD}(G, G') \leq ID_\xi(G, G') \leq \text{RD}(G, G'). \quad (5.9) \]

The intermediate distance induces a metric (called the intermediate metric); it is automatically positively semidefinite, symmetric, and the triangle inequality is satisfied (cf. Theorems 3.4 and 5.5).

5.1. Evaluation of reaction distance

The chemical distance between two isomeric graphs was defined with the help of their maximal common subgraph, this definition makes possible, in principle, the evaluation of chemical distances for an arbitrary pair of isomeric graphs. The reaction distance is defined as the graph distance in \(\mathcal{F}_{pq}\). It means that in order to evaluate the reaction distance we have to know this graph \(\mathcal{F}_{pq}\) for the family \(\mathcal{F}_{pq}\). Unfortunately, its explicit construction is possible only for relatively small families \(\mathcal{F}_{pq}\), i.e., for smaller values of integers \(p\) and \(q\). For greater values of \(p\) and \(q\) (e.g. for \(pq \geq 10\)) a construction of the graph \(\mathcal{F}_{pq}\) is a hopeless task. Therefore it is very important to have an approach which does not require to know the whole graph \(\mathcal{F}_{pq}\) but only some part of it (a subgraph) which still contains all the minimal paths for a preselected pair of graphs from \(\mathcal{F}_{pq}\).

We turn our attention on the so-called bilateral approach [18]. Although it leads to a combinatorial explosion of the produced intermediate graphs, it can serve as a model prototype of algorithms for the evaluation of reaction distances. Let \(G_1\) and \(G_2\) be a pair of graphs from the family \(\mathcal{F}_{pq}\). If these graphs are isomorphic, then their reaction distance is zero, \(\text{RD}(G_1, G_2) = 0\) (cf. equation (5.4a)). Therefore we shall assume that \(G_1\) and \(G_2\) are nonisomorphic, i.e., \(\text{RD}(G_1, G_2) > 0\). 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 \(\xi = \alpha, \beta\), new left set \(\mathcal{L}_1\) is formed from all nonisomorphic resulting product graphs. In the \(k\)th step (if \(k \geq 2\)) the left set \(\mathcal{L}_k\) consists of those graphs not \(\mathcal{L}_{k-2}\) that are products of elementary transformations of graph in \(\mathcal{L}_{k-1}\). In a completely similar way the right sets \(\mathcal{R}_0, \mathcal{R}_1, \ldots, \mathcal{R}_f\) are constructed. 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 mutually isomorphic, formally $\mathcal{L}_k \cap \mathcal{R}_l \neq \emptyset$, then the reaction distance is $\text{RD}(G_1, G_2) = k + l$, in the opposite case we continue the process of construction of new left and right sets of graphs. The outlined bilateral approach for the evaluation of the reaction distance is summarized in the form of an algorithm.

**Algorithm 5.11.**

1. **Step 1.** (Initialization). Input of graphs $G_1$ and $G_2$;
   
   - $k := l := 0$; $\mathcal{L}_0 := \{G_1\}$; $\mathcal{R}_0 := \{G_2\}$;

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 $\mathcal{L}_k := \mathcal{L}_k \setminus \mathcal{L}_{k-2}$;
   
   - if $\mathcal{L}_k \cap \mathcal{R}_l \neq \emptyset$ then go to Step 4;

3. **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 $\mathcal{R}_l := \mathcal{R}_l \setminus \mathcal{R}_{l-2}$;

   - if $\mathcal{L}_k \cap \mathcal{R}_l \neq \emptyset$ then go to Step 4 else go to Step 2;

4. **Step 4.** $\text{RD}(G_1, G_2) = k + l$;

5. **Step 5.** End of algorithm;

**Example 5.12.** Apply Algorithm 5.11 for the evaluation of the reaction distance between graphs $G_1$ and $G_2$ from Example 3.1.

1. **Step 1.** $k := 0$, $l := 0$; $\mathcal{L}_0 := \{G_1\}$, $\mathcal{R}_0 := \{G_3\}$.

2. **Step 2.** $k := 1$, $\mathcal{L}_1 := \{G_2\}$, $\mathcal{L}_1 \cap \mathcal{R}_0 = \emptyset$.

3. **Step 3.** $l := 1$. $\mathcal{R}_1 := \{G_3\}$, $\mathcal{L}_1 \cap \mathcal{R}_1 = \emptyset$.

4. **Step 2.** $k := 2$, $\mathcal{L}_2 := \{G_3, G_6\}$, $\mathcal{L}_2 \cap \mathcal{R}_1 = \{G_3\}$.

5. **Step 4.** $\text{RD}(G_1, G_3) = 2 + 1 = 3$.

Schematically the above successive steps are illustrated by
The main drawback of the bilateral approach is that its numerical efficiency fast decreases (a combinatorial explosion) for pairs of graphs with greater reaction distance. Therefore, for actual applications we have to modify the algorithm by the following very efficient heuristic: The set $\mathcal{L}_k$ is composed of only those graphs for which a chemical distance between them and the graph $G_2$ is smaller than or equal to a chemical distance between their predecessors (from the set $\mathcal{L}_{k-1}$) and the graph $G_2$. The same rule should also hold for the graphs from the set $\mathcal{R}_l$. This heuristic substantially reduces the cardinality of the graph sets $\mathcal{L}_k$ and $\mathcal{R}_l$, the graphs from the "blind" ways which do not connect the graphs $G_1$ and $G_2$ by a path are removed. Unfortunately, since the above rule is only a heuristic, the produced reaction distance between $G_1$ and $G_2$ need not be a minimal one but in some special cases it might be greater than its optimal (i.e., minimal) value. Therefore, the reaction distance evaluated by the bilateral approach accelerated by the above considered heuristic is an upper estimate of its exact value.

Recently, Koča [16,17] has suggested an effective method for an upper estimate of the reaction distance between two graphs $G_1$ and $G_2$ based on the so-called minimal covering of the reaction graph $G_R$ assigned to the transformation $G_1 \rightarrow G_2$. This transformation may be "algebraicized" by making use of the reaction graph (see equation (4.5)),

$$G_1 + G_R = G_2 \approx G_2.$$  \hspace{1cm} (5.10)

We try to decompose the reaction graph $G_R$ into a sum of elementary reaction graphs (5.3a-b) composed of a minimal number of terms,

$$G_R = \sum_{i=1}^{l} G_R^{(i)}. \hspace{1cm} (5.11)$$

The minimal value of $l$ corresponds to an upper bound of the reaction distance (see Definition 5.3). All admissible ways of decomposition of $G_R$ into $l$ elementary reaction graphs determine "shortest" paths connecting the graphs $G$ and $G'$ in $\mathcal{G}^{RD}$. We see that this approach may be understood as a specially "controlled" realization of the bilateral algorithm accelerated by the heuristic of accounting for only those graphs with smaller chemical distances.

6. Concluding remarks

Two different graph metrics were introduced for purposes of the graph theory
model of organic chemistry. Both these metrics make it possible to quantify a similarity (or dissimilarity) between graphs taken from the same family \( S_{pq} \) of isomeric graphs. Although the introduced metrics are not "isometric" they are not fully independent. We have proved that the graph of reaction distances \( D_{pq}^{RD} \) is a subgraph of the graph of chemical distances \( D_{pq}^{CD} \), see equation (5.5). This very important property implies that the chemical and reaction distances are restricted by the inequality (5.6). A convex linear combination of the distances provides the so-called intermediate distance which is also a metric over the family \( S_{pq} \). Our preliminary studies indicate that the chemical distance might be of value also outside of the graph theory model of organic chemistry. For instance, the problem of reconstruction [6] of a graph from its subgraphs that are generated from the original one by a deletion of one vertex can be substantially forced towards to unexpected looks at the problem and new partial results. This work is in progress and will be published elsewhere.

References