

НЕКОТОРЫЕ ХАРАКТЕРИСТИКИ И ОПИСАНИЕ
МОЛЕКУЛЯРНЫХ (ПОМЕЧЕННЫХ) ГРАФОВ И ИХ ПОДГРАФОВ

Л.И. Макаров

Компьютерный анализ зависимости "структура - свойство" химических соединений основан на гипотезе о том, что свойства соединений определяются соответствующими структурными фрагментами и структурно подобные химические соединения обладают близкими свойствами [1]. Целью анализа является поиск структурных фрагментов, определяющих свойства соединений, и прогнозирование свойств других соединений на основе данных о вхождении в них таких фрагментов.

Для компьютерной обработки данных о молекулярных структурах и их фрагментах необходимо ввести соответствующие математические модели. Обычно, в качестве их моделей используются молекулярные графы и их подграфы. При этом в зависимости от конкретных особенностей исследуемого семейства химических соединений для их описания могут быть использованы подграфы разных видов. Например, подграфы связанные или несвязанные, порожденные или частичные; подграфы, общие для графов нескольких соединений, наибольшие общие по порядку или определенного порядка с наибольшим числом ребер. Кроме того, для описания структур используют и количественные их характеристики - метрические характеристики, топологические и информационные индексы, оценки взаимного положения подграфов, оценки степени подобия графов и др.

В данной работе дано систематизированное описание подграфов, относящихся к выделенному подграфу данного графа, приведены некоторые утверждения о свойствах общих подграфов пары графов, для взвешенных графов описаны количественные характеристики взаимного положения подграфов в графе и оценки степени подобия графов, рассмотрены способы описания молекулярных (помеченных) графов, использующие их характеристики и подграфы.

Молекулярные графы и их подграфы

В качестве моделей молекулярных структур химических соединений и их фрагментов обычно используются молекулярные графы и их подграфы. Молекулярным графом химического соединения называют помеченный граф, вершины которого соответствуют атомам или группам атомов химического соединения, а ребра - химическим связям между его атомами. При этом каждому фрагменту молекулярной структуры соединения соответствует определенный подграф молекулярного графа. Метки графа и его элементов (вершин и ребер) могут содержать различные данные о молекулярной структуре, в том числе данные о типах элементов структуры и значениях их электронно-топологических параметров, например, для молекулы - имя химического соединения, объем и дипольный момент молекулы, энергии граничных молекулярных орбиталей; для атомов - имя, эффективный заряд, заселенность высшей занятой и нижней свободной молекулярных орбиталей; для межатомных связей - индекс Уайберга (кратность) и пространственная длина связи.

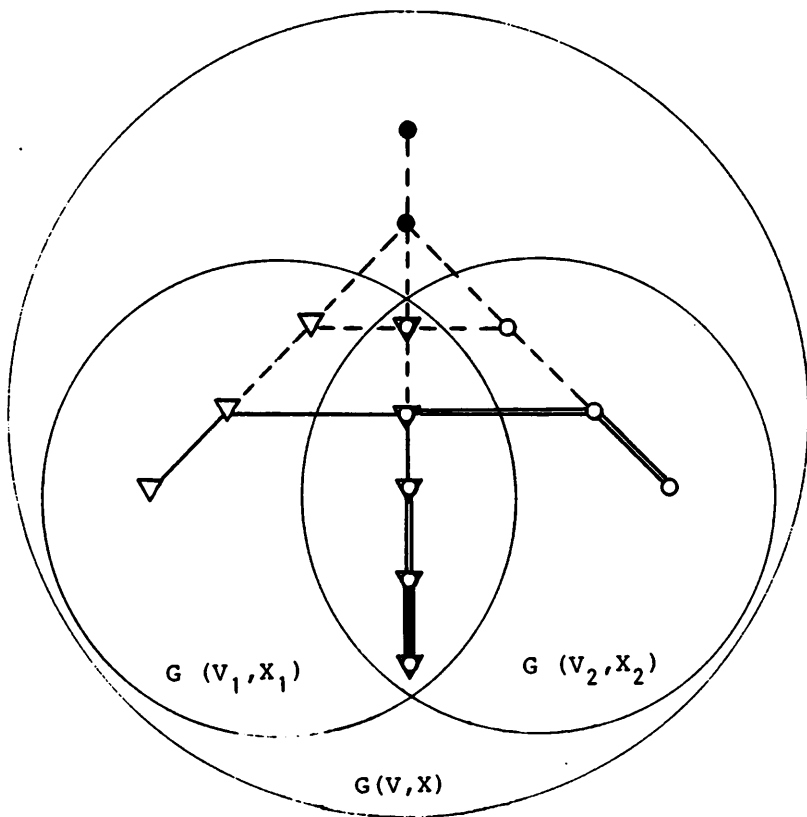
Приведенные ниже определения типов подграфов в основном соответствуют определениям из [2,3] и для простоты описания относятся к непомеченным графам и их подграфам.

Пусть $G(V, X)$ – граф, имеющий множество вершин $V = \{v_i\}$, $i = 1, 2, \dots, p_G$, и множество ребер $X = \{x_j\}$, $j = 1, 2, \dots, q_G$. Каждое ребро $x \in X$ соответствует паре (v, u) вершин $v, u \in V$, при этом ребро $x = (v, u)$ называют инцидентным вершинам v и u , а вершины v и u , соединяемые ребром x , называются смежными. Граф $G_1(V_1, X_1)$ называют подграфом (частью) графа G , $G_1 \subseteq G$, если его множества V_1 и X_1 содержатся в соответствующих множествах V и X , $V_1 \subseteq V$, $X_1 \subseteq X$. Каждый подграф $G_1 \subseteq G$ однозначно задается (порождается) множеством своих ребер $X_1 \subseteq X$ и имеет множество $V(X_1) = V_1 \subseteq V$ всех вершин, соединяемых этими ребрами (изолированные вершины могут быть заданы, если для каждой вершины $v \in V$ графа G ввести ребро-петлю $x = (v, v) \in X$). В этом случае подграф G_1 будем называть частичным (или реберно-порожденным) и обозначать как $G_1 = G(V(X_1), X_1) = \langle X_1 \rangle$. Подграфом G_1^V , порожденным (или вершинно-порожденным) в графе G множеством вершин $V_1 \subseteq V$, называют подграф $G_1^V = G(V_1, X(V_1))$, в котором множество ребер $X_1 = X(V_1) \subseteq X$ состоит из всех ребер соединяющих в G вершины из V_1 , и обозначают $G_1^V = \langle V_1 \rangle_G$. Очевидно, что $G_1^V = \langle X(V_1) \rangle$.

Из множеств A и B элементов графа с помощью известных операций могут быть образованы другие множества: объединение $A \cup B$, пересечение $A \cap B$, разность $A \setminus B$; если $B \subseteq A$, то $A \setminus B$ называют дополнением B в A .

Пусть в графе $G(V, X)$ заданы подграфы $G_1 = G(V_1, X_1)$ и $G_2 = G(V_2, X_2)$. Введем обозначения (рис.1): $V_0 = V \setminus (V_1 \cup V_2)$, $X_0 = X \setminus (X_1 \cup X_2)$, $V_3 = V_1 \cap V_2$, $X_3 = X_1 \cap X_2$;

$X_1^{jk} = \{x = (v, u) : x \in X_1, v \in V_j, u \in V_k\}$ – множество всех ребер $x \in X_1$, соединяющих вершину v из V_j с вершиной u из V_k ; $i, j, k \in \{0, 1, 2, 3\}$;



Обозначения

∇ $v_1 \setminus v_3$

\circ $v_2 \setminus v_3$

∇ $v_3 = v_1 \cap v_2$

\bullet $v_0 = v \setminus (v_1 \cup v_2)$

— x_1

== x_2

=== $x_3 = x_1 \cap x_2$

--- $x_0 = x \setminus (x_1 \cup x_2)$

Рис. 1. Подграфы графа G

$X_G^V = \{x = (v, u) : x \in X', u \in V'\}$ - множество всех ребер, инцидентных в графе $G' = G(V', X') \subseteq G$ вершине $v \in V'$.

Отсюда следует $X_1 = X_1^{11}$, $X_2 = X_2^{22}$, $X_o = \bigcup_{j,k} X_o^{jk}$, $X_i^{jk} = X_i^{kj}$, $X_i^{i3} \subseteq X_i$, $X_i^{33} \subseteq X_i$; $X_3 = X_1^{33} \cap X_2^{33}$; при $i \neq 0$ в X_i^{jk} индексы $j, k \in \{i, 3\}$, т.е. $X_1^{12} = X_1^{22} = X_1^{23} = X_2^{11} = X_2^{12} = X_2^{13} = \emptyset$.

Множества элементов графа, полученные в результате применения операций объединения, пересечения и разности, порождают соответствующие подграфы графа G .

Суммой $G_{12}^U = G_1 \cup G_2$, пересечением $G_{12}^\cap = G_1 \cap G_2$ и разностью $G_{12} = G_1 \setminus G_2$ двух подграфов $G_1 \subseteq G$ и $G_2 \subseteq G$ называют соответственно подграфы $G_{12}^U = \langle X_1 \cup X_2 \rangle$, $G_{12}^\cap = \langle X_1 \cap X_2 \rangle$ и $\bar{G}_{12} = \langle X_1 \setminus X_2 \rangle$. Звездой S_G^V , вершины $v \in V'$ в графе $G' \subseteq G$ называют подграф $S_G^V = \langle X_G^V \rangle$.

Определим некоторые вспомогательные подграфы, относящиеся к подграфу $G_1 \subseteq G$ при отсутствии подграфа G_2 , т.е. $G_2 = \emptyset$, $V_2 = \emptyset$, $X_2 = \emptyset$ и $V_o = V \setminus V_1$, $X_o = X \setminus X_1$, $V_3 = \emptyset$, $X_3 = \emptyset$, $i, j, k \in \{0, 1\}$:

$$\bar{G}_1 = G \setminus G_1 = \langle X_o \rangle - \text{дополнение } G_1.$$

$$G_o^V = \langle V_o \rangle_G = \langle X_o^{oo} \rangle - \text{вершинное дополнение } G_1,$$

$$G_1^L = \langle X_o^{11} \rangle \cup \langle X_o^{01} \rangle - \text{граница } G_1,$$

$$G_1^N = G_1 \cup G_1^L - \text{окрестность } G_1.$$

Легко видеть, что $G_1 = G(V_1, X_1) = \langle X_1 \rangle \subseteq G_1^V = \langle V_1 \rangle_G$;
 $S_{G_1}^V \subseteq S_G^V$; $G_1 \cup \bar{G}_1 = G$; $\bar{G}_1 = G_o^V \cup G_1^L$; $G_1^V = \langle V_1 \rangle_G = G_1 \cup \langle X_o^{11} \rangle$;
 $G_1^N = G \setminus G_o^V = \bigcup S_G^V$; $S_G^V = G_1^N$ при $G_1 = v$.

Аналогичные понятия можно ввести при рассмотрении двух подграфов G_1 и G_2 графа G , например

$G_{12}^{\Delta} = G_1 \Delta G_2 = \langle (x_1 \cup x_2) \setminus x_3 \rangle$ - симметрическая разность G_1 и G_2 ;

$\bar{G}_{12}^V = \langle v_1 \setminus v_2 \rangle = \langle x_1 \setminus (x_1^{13} \cup x_1^{33}) \rangle$ - вершинное дополнение G_2 в G_1 .

$G_{12}^L = \langle x_1^{13} \rangle \cup \langle (x_1^{33} \setminus x_3) \rangle$ - граница G_2 в G_1 ,

$G_{12}^N = \langle v_3 \rangle \cup G_{12}^L = \langle x_2^{33} \cup x_1^{33} \cup x_1^{13} \rangle$ - окрестность G_2 в G_1 .

Очевидно, что $G_{12}^{\Delta} = \bar{G}_{12} \cup \bar{G}_{21}$ и $G_{12}^L \neq G_{21}^L = \langle x_2^{23} \rangle \cup \langle (x_2^{33} \setminus x_3) \rangle$.

Рассмотрим далее подграфы в разных графах. Два помеченных графа G и H называют изоморфными ($G \simeq H$), если между их множествами вершин V_G и V_H существует взаимно-однозначное соответствие, сохраняющее метки вершин и их смежность по соответствующим помеченным ребрам. Соответствие V_G и V_H задают в виде подстановки $\tau: p(V_G) \rightarrow p(V_H)$ множества номеров вершин $p(V_G)$ во множество номеров вершин $p(V_H)$.

Граф F называют общим подграфом графов G и H , если в них существуют подграфы $G' \subseteq G$ и $H' \subseteq H$, изоморфные графу F , $G' \simeq F \simeq H'$. Максимальным по включению порожденным (частичным) общим подграфом назовем такой подграф, что в G и H не существует порожденного (частичного) общего подграфа, строго его содержащего.

Пересечением Q графов G и H называют наибольший по числу вершин общий подграф графов G и H , изоморфный порожденным подграфам $G' \subseteq G$ и $H' \subseteq H$. Перекрытием C назовем наибольший по числу ребер общий частичный подграф G и H . Наибольшее по числу ребер пересечение графов G и H обозначим через \hat{Q} , а наибольшее по числу неизолированных вершин перекрытие - через \hat{C} .

Для иллюстрации свойств общих подграфов пары графов G и H приведем следующие утверждения.

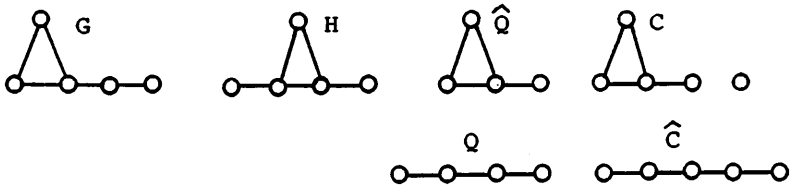


Рис. 2

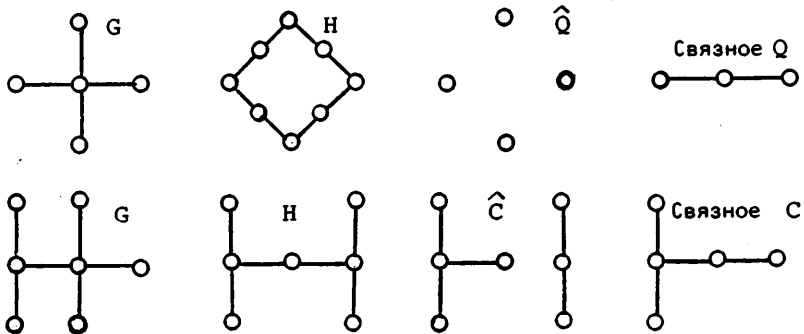


Рис. 3

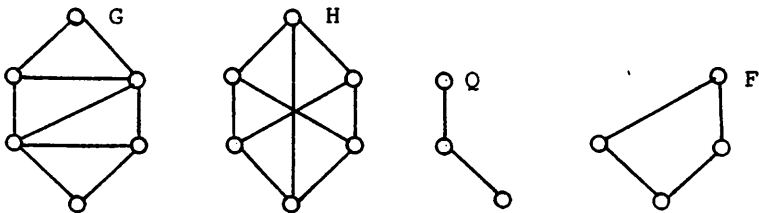


Рис. 4

1. Пара графов G и H может иметь несколько пересечений Q и перекрытий C (рис.2). При этом каждое пересечение или перекрытие может соответствовать разным, хотя и изоморфным подграфам одного графа.

2. Наибольшее пересечение \hat{Q} (перекрытие \hat{C}) пары графов может не содержать их связное пересечение (перекрытие) (рис.3).

3. Пересечение Q (рис.4) и перекрытие C (рис.5) пары графов могут не содержать их общий подграф F . Для помеченных графов, соответствующие непомеченные графы которых изоморфны, это утверждение также справедливо (рис.6).

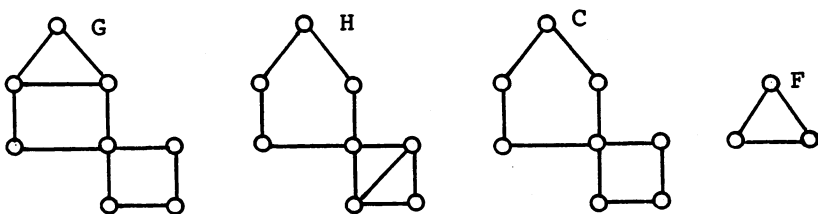


Рис. 5

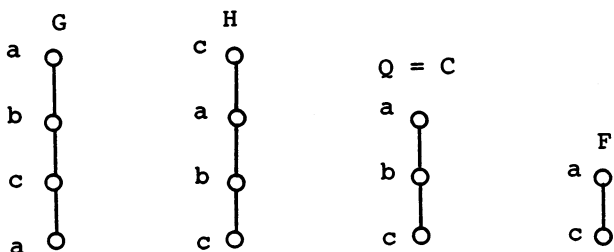


Рис. 6

4. Для любого общего подграфа пары графов существует содержащий его подграф из множества всех максимальных по включению частичных общих подграфов.

Все приведенные выше определения естественным образом могут быть обобщены для случая помеченных графов. В помеченных графах метки вершин и ребер могут быть составными, т.е. кроме собственных меток могут содержать и собственные метки инцидентных и смежных элементов. В этом случае операции удаления подграфа; замены подграфа одной вершиной, метка которой содержит все данные о подграфе, и обратные им операции могут быть произведены однозначно. При установлении изоморфизма графов метки их элементов могут учитываться как полностью, так и частично.

Гиперграфовые модели молекулярных структур могут быть сведены к молекулярным графам, поскольку любой гиперграф имеет представление в виде помеченного графа [5].

Количественные характеристики графов и их подграфов

Определим некоторые количественные характеристики графа и его подграфов. Для обыкновенного (без кратных ребер и петель) графа $G(V, X)$, $|V| = p_G$, $|X| = q_G$, расстоянием $d_G(v, u)$ между его вершинами $v, u \in V$ называют число ребер кратчайшей простой цепи, соединяющей их в графе G . В связном обыкновенном графе расстояние удовлетворяет аксиомам метрики, поэтому количественные характеристики такого графа и его подграфов, определяемые на основе понятия расстояния, называют метрическими. К их числу относятся эксцентриситет графа $e(G) = \sum e(v)$ по $v \in V$, $e(v) = \max d_G(v, u)$ по $u \in V$; дистанция графа (индекс Винера) $D(G) = \frac{1}{2} \sum d_G(v, u)$, $v, u \in V$; центрический индекс $V_C = \sum d_G^2(v, v_C)$, где вершина v_C принадлежит центру графа, т.е. множеству вершин, для которых $e(v_C) = \min e(v)$, $v \in V$; и другие [4, 6].

Кроме метрических для описания графа используют и другие количественные характеристики [6], например, топологические индексы - индекс молекулярной связности Рандича $\chi = \sum \sqrt{s_v s_u}$, по

$x = (v, u) \in X$, s_v - степень вершины v ; информационные топологические индексы вида $I = -\sum p_t \log_2 p_t$, $t = 1, 2, \dots, n$; p_t - вероятность вхождения в молекулярный граф t -го фрагмента из заданного набора фрагментов и т.д.

Количественные характеристики, аналогичные характеристикам обыкновенного графа, могут быть введены и для взвешенного графа. Граф $G(V, X)$ называют взвешенным, если каждому его ребру $x = (v, u) \in X$, $v, u \in V$, приписано положительное действительное число $l_G(x)$ - длина (вес) ребра. Взвешенный граф G задают взвешенной матрицей смежности $A(G) = \| a_{ij} \|$, $i, j = 1, 2, \dots, p_G$, где $a_{ij} = l_G(x_{ij})$ при $x_{ij} = (v_i, v_j) \in X$ и $a_{ij} = 0$ при $x_{ij} \notin X$. Каждому взвешенному графу можно поставить в соответствие обыкновенный граф с матрицей смежности $A(G')$, если принять $a'_{ij} = 1$ для всех $x_{ij} \in X$, $i \neq j$, $a'_{ii} = 0$ и $a'_{ij} = 0$ при $x_{ij} \notin X$.

Для связного подграфа $G' = (V', X') \subseteq G$ графа G определим его длину как величину $l(G') = \sum l_G(x)$ по $x \in X'$. Расстоянием $r_G(v, u)$ (или $r_{G'}(v, u)$) между вершинами $v, u \in V$ (или $v, u \in V'$) назовем длину кратчайшей цепи, соединяющей вершины v и u в графе G (или в его подграфе G'). В общем случае величина $r_G(v, u)$ не является метрикой и $r_G(v, u) \geq r_{G'}(v, u)$.

Тогда для каждой метрической характеристики обыкновенного графа можно ввести аналогичную характеристику для взвешенного графа заменой метрики $d_G(u, v)$ на расстояние $r_G(v, u)$, а степени s_v вершины на величину суммы длин, инцидентных ей ребер, $l(S_G^V)$.

Количественные характеристики можно использовать и для оценки взаимного положения подграфов F и H в графе G (в частности один из подграфов может совпадать с графом G). Пусть для некоторого графа $F \subseteq G$ в графе G существует набор изоморфных подграфов F_i , $i = 1, 2, \dots, t$, $F_i \cong F$, с попарно несовпадающи-

ми множествами номеров вершин $n(V_{F_i})$. Тогда будем говорить, что в граф G подграф F входит в разных (с точностью до симметрий) положениях, которые задаются подстановками номеров вершин $\tau_i: n(V_F) \rightarrow n(V_{F_i})$.

В [7] для оценки заданного взаимного положения подграфов F и H в обыкновенном графе G предлагается величина нормированной дистанции между подграфами: $\frac{1}{D(G)} \sum \sum d_G(v, u)$, $v \in V_F$, $u \in V_H$. Эта величина при $F \subseteq H$ зависит от отклонения подграфа F от "центра" H , но при совпадении в G изоморфных подграфов F и H она отлична от нуля.

Определим здесь некоторые характеристики, позволяющие количественно оценивать заданное взаимное положение подграфов F и H взвешенного графа G :

$$\text{расстояние } r(F, H) = \min_v r_G(v, H) = \min_v (\min_u r_G(v, u)),$$

$$v \in V_F, \quad u \in V_H;$$

$$\text{удаление } R(F, H) = R(F) + R(H) = \sum_v r_G(v, H) + \sum_u r_G(u, F),$$

$$v \in V_F, \quad u \in V_H;$$

$$\text{дистанционное удаление } L(F, H) = D_G(F, H) - D_G(F \cap H),$$

где $D_G(F, H) = \sum \sum r_G(v, u)$, $v \in V_F$, $u \in V_H$ - дистанция между F и H в G и $D_G(F \cap H) = \sum \sum r_G(v, u)$, $v, u \in V_{F \cap H}$ - дистанция подграфа $F \cap H$ в G ; и аналогичные нормированные по порядку $F \cup H$ величины:

$$\rho(F, H) = \frac{1}{P_{F \cup H}} R(F, H),$$

$$\lambda(F, H) = \frac{1}{P_{F \cup H}} L(F, H).$$

Расстояние $r(F, H) = 0$ при $V_F \cap V_H \neq \emptyset$. Остальные величины при совпадении в G множеств вершин $V_F = V_H$ подграфов обращаются в нуль, однако при этом подграфы F и H могут иметь несовпадающие множества ребер. Для учета различия множеств ребер подграфов обыкновенного графа при $V_F \cap V_H \neq \emptyset$ можно использовать порядок симметрической разности множеств ребер $X_F \Delta X_H$: $q_{F\Delta H} = q_F + q_H - 2q_{F\cap H}$, и для оценки взаимного положения подграфов рассматривать величины $R'(F, H) = R(F, H) + q_{F\Delta H}$, $L'(F, H) = L(F, H) + q_{F\Delta H}$ и нормированные по $q_{F\cup H}$: $\rho'(F, H) = \rho(F, H) + q_{F\Delta H}/q_{F\cup H}$, $\lambda'(F, H) = \lambda(F, H) + q_{F\Delta H}/q_{F\cup H}$.

Для взвешенного графа G и его подграфов F и H в оценки их взаимного положения нужно ввести величины $1(\langle X_F \Delta X_H \rangle)$ и $1(\langle X_F \cup X_H \rangle)$, аналогичные $q_{F\Delta H}$ и $q_{F\cup H}$.

Для случая, когда подграфы содержат хотя бы одну общую вершину, $V_F \cap V_H \neq \emptyset$, в приведенных характеристиках расстояния $r(v, u)$ между парой вершин v и u могут быть определены как по цепям графа G , так и по цепям графа $F \cup H$, т.е. для вычисления значений этих характеристик могут быть использованы величины $r_G(v, u)$ или $r_{F\cup H}(v, u)$ в зависимости от конкретных особенностей решаемой задачи.

Степень подобия графов также можно оценивать количественными характеристиками. Например, для обыкновенных графов - с помощью величин расстояний между графами G и H [8]:

$$p(G, H) = p_G + p_H - 2p_{G\cap H}, \quad q(G, H) = q_G + q_H - 2q_{G\cap H}$$

или соответствующих величин, нормированных по $(p_G + p_H)$ и $(q_G + q_H)$, где $q_{G\cap H}$ - число ребер, а $p_{G\cap H}$ - порядок наибольшего общего подграфа графов G и H .

Для оценки степени подобия взвешенных графов вначале определим расстояние между симметрическими матрицами A и B с дей-

ствительными элементами a_{ij}, b_{ij} , $i, j = 1, 2, \dots, k$, как величину $N(A, B) = \min_{\tau} \sum_{i=1}^k \sum_{j=1}^k |a_{ij} - b_{\tau(i)\tau(j)}|$ по всем τ - одновременным перестановкам строк и соответствующих столбцов матрицы B .

Тогда степень подобия - расстояние $S(G, H)$ между взвешенными графами G и H , $p_G = p_H$, можно оценивать расстоянием между их взвешенными матрицами смежности $A(G)$ и $A(H)$: $S(G, H) = N(A(G), A(H))$. При $p_G < p_H$ в граф G вводится $(p_G - p_H)$ изолированных вершин, а матрица $A(G)$ добавлением элементов $a_{kl} = 0$, $k, l = p_G + 1, \dots, p_H$, расширяется до порядка p_H .

При заданной перестановке τ множества ребер $x = (v_i, v_j) \in X_G$ и соответственно $\tau(x) = (v_{\tau(i)}, v_{\tau(j)}) \in X_H$, таких, что $l_G(x) \neq 0$ и $l_H(\tau(x)) \neq 0$ одновременно, обозначим через X_G^{τ} и X_H^{τ} . В обыкновенных графах, соответствующих графам G и H , эти множества ребер порождают подграфы, изоморфные их общему подграфу.

Тогда расстояние между графами можно представить в виде

$$S(G, H) = \min_{\tau} [1(\langle X_G \setminus X_G^{\tau} \rangle) + 1(\langle X_H \setminus X_H^{\tau} \rangle) + \sum_x |l_G(x) - l_H(\tau(x))|]$$

по $x \in X_G^{\tau}$ и всем перестановкам τ .

Для практического применения может оказаться полезной величина расстояния, получающаяся из $S(G, H)$, если ограничиться перестановками, множества ребер X_G^{τ} и X_H^{τ} которых порождают наибольшие общие подграфы соответствующих обыкновенных графов, с учетом симметрий этих подграфов.

Описание молекулярных графов

Таким образом, введенные понятия подграфов молекулярных графов и их количественных характеристик позволяют достаточно точно описывать фрагментный состав и структуру молекул исследуемого множества химических соединений [4,9]. Квантово-химические, геометрические, топологические, физико-химические, биологические и другие свойства молекул могут быть получены расчетным или экспериментальным путем и заданы в виде значений соответствующих параметров.

Для выбранного множества (выборки) соединений $S = \{s_i\}$, $i = 1, 2, \dots, n$, описание их молекулярных графов с помощью количественных характеристик (параметров) задают в векторном виде, т.е. каждому молекулярному графу G_i ставят в соответствие вектор (набор) $t_i = (t_{i1}, \dots, t_{ij}, \dots, t_{im})$, j -я компонента которого равна значению j -го параметра из множества параметров $h = \{h_j\}$, $j = 1, 2, \dots, m$, выбранного для описания молекулярных графов. Так же, в векторном виде задают описание фрагментарного состава молекулярного графа, при этом значением j -й компоненты вектора t_i является число t_{ij} различных вхождений в граф G_i фрагмента F_j из множества фрагментов $F = \{F_j\}$, $j = 1, 2, \dots, m$, или значение средней оценки положения F_j в G_i . В качестве используемого множества фрагментов F можно выбрать некоторый их набор из известного множества фрагментов, соответствующего семейству соединений, которому принадлежит исследуемая выборка. Например, набор фрагментов, содержащий отдельные атомы соединений, связи пар атомов, окрестности атомов, циклы заданной длины, гетероцепи, характерные функциональные группы химических структур и т.д. Кроме того, используемое множество фрагментов F может содержать заранее неизвестные фрагменты, которые входят в соединения конкретной выборки S , например; мно-

жество всех фрагментов соединений выборки, множество наибольших попарно общих фрагментов соединений и т.д.

При векторном задании молекулярных графов описание исследуемого множества соединений - выборки S формируется в виде таблицы (матрицы) T , в которой i -я строка является векторным описанием t_i соединения s_i , $i = 1, 2, \dots, n$, а j -й столбец соответствует j -му параметру. В задачах распознавания образов, обнаружения эмпирических закономерностей такие таблицы называют "таблицы объект-признак", поскольку в них объекты - химические соединения заданы значениями их признаков - параметров. При этом разные параметры в описании соединений могут относиться к разным типам измерительных шкал - количественной, порядковой или шкале наименований (номинальной) [10]. Например, физико-химические параметры молекулы и индексы графов относятся к количественной шкале измерения, токсичность веществ может измеряться в порядковой шкале, а признак наличия или отсутствия данного фрагмента в соединении относится к номинальной шкале измерения. При этом значения некоторых признаков могут относиться к разным шкалам. Например, признак положения фрагмента, входящего в некоторый молекулярный граф, измеряется в количественной шкале, а при отсутствии фрагмента в данном соединении этот признак принимает значение из номинальной шкалы.

Описание структуры молекулы химического соединения задается с помощью матрицы смежности $A(G)$ его молекулярного графа G . Элементом a_{ij} матрицы $A(G) = \|a_{ij}\|$ молекулярного графа $G(V, X)$ порядка p является метка ребра $(v_i, v_j) \in X$, $i, j = 1, 2, \dots, p$, например, кратность и/или длина межатомной связи, а элементом a_{ii} - метка вершины $v_i \in V$.

Если известны набор и положение фрагментов $\tilde{F} = \{F_k\}$, $k = 1, 2, \dots, n$, входящих в молекулярный граф G химического соединения, и $G = \cup F_k$, то структуру его молекулы можно задать с по-

мощью матрицы смежности $A(G_F)$ графа фрагментов $G_F(\tilde{F}, X_F)$. Вершины графа фрагментов соответствуют фрагментам $F_k(V_k, X_k)$, а ребра - связям между фрагментами, т.е. элемент a_{k1} матрицы $A(G_F)$, $k, 1 = 1, 2, \dots, n$, соответствует границе F_{k1}^L фрагмента F_k в F_1 и может быть задан в виде множества ребер X_{k1}^L или в виде матрицы смежности $A(F_{k1}^L)$, $a_{k1} \neq a_{1k}$, а элемент a_{kk} - в виде множества ребер X_k или в виде матрицы смежности $A(F_k)$.

Взаимное положение фрагментов из некоторого их набора \tilde{F} в молекулярном графе G может быть описано с помощью матрицы $A(\tilde{F})$, элемент a_{k1} которой равен количественному значению выбранной оценки относительного положения фрагментов F_k и F_1 в графе G , $k, 1 = 1, 2, \dots, n$, если $F_k \subseteq G$ и $F_1 \subseteq G$, и a_{k1} равен номинальному символу отсутствия фрагмента, если F_k или F_1 не входят в G .

Количественные оценки попарного подобия в семействе графов $\{G_i\}$, $i = 1, 2, \dots, n$, также задают в виде матрицы $A(\tilde{G})$ смежности взвешенного графа подобия $\tilde{G}(V, X)$, вершины которого соответствуют графам семейства, а элемент a_{ij} матрицы $A(\tilde{G})$ равен значению оценки степени подобия графов G_i и G_j , $i, j = 1, 2, \dots, n$ [8].

З а к л ю ч е н и е

Рассмотренные способы описания и количественные характеристики молекулярных графов и их фрагментов позволяют для исследуемого семейства химических соединений выбрать описание подходящее по степени детализации и адекватности реальным структурам. Некоторые из способов описания и характеристик молекулярных графов и их фрагментов нашли применение в компьютерных системах структурно-метрического анализа и классификации моле-

кулярных графов СИСТРАН, САКМГ, МЕТАХИМ и др., разработанных в Институте математики СО РАН.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 93-03-18657).

Л и т е р а т у р а

1. СТЬЮПЕР Э., БРЮГГЕР У., ДЖУРС П. Машинный анализ связи химической структуры и биологической активности. - М.: Мир, 1982.

2. ОРЕ О. Теория графов. - М.: Наука, 1980.

3. ТАТТ У. Теория графов. - М.: Мир, 1988.

4. МАКАРОВ Л.И., СКОРОБОГАТОВ В.А. Комплекс программ для исследования зависимости "структура-свойство" химических соединений //Алгоритмический анализ графов и его применения. - Новосибирск, 1988. - Вып. 127: Вычислительные системы. -С.92-127.

5. ЗЫКОВ А.А. Гиперграфы //Успехи мат.наук. - 1974.-Т.29, вып. 6. - С. 89-154.

6. СТАНКЕВИЧ М.И., СТАНКЕВИЧ И.В., ЗЕФИРОВ Н.С. Топологические индексы в органической химии //Успехи химии. - 1988. - Т. 57. - С. 337-366.

7. ДОБРЫНИН А.А., СКОРОБОГАТОВ В.А. Метрические инварианты подграфов молекулярных графов //Математические методы в химической информатике. - Новосибирск, 1991. - Вып. 140: Вычислительные системы. - С. 3-62.

8. МАКАРОВ Л.И. Алгоритмы таксономии вершин взвешенного графа //Там же. - С. 130-142.

9. РОЗЕНБЛИТ А.Б., ГОЛЕНДЕР В.Е. Логико-комбинаторные методы конструирования лекарств. - Рига: Зинатне, 1984.

10. ЗАГОРУЙКО Н.Г., ЕЛКИНА В.Н., ЛБОВ Г.С. Алгоритмы обнаружения эмпирических закономерностей. - Новосибирск: Наука, 1985.

Поступила в ред.-изд.отд.

23 октября 1993 года