АНАЛИЗ ДАННЫХ В ЭКСПЕРТНЫХ СИСТЕМАХ
(Вычислительные системы)

1986 год Выпуск 117

УДК 519.17:5/6

ВЛИЯНИЕ СТРУКТУРНЫХ ПРЕОБРАЗОВАНИЙ ГРАФА НА ЗНАЧЕНИЕ ЕГО ДИСТАНЦИИ

В.А.Скоробогатов, А.А.Добрынин

Теория графов является необходимым математическим средством для анализа и классификации различных классов соединений по их структурным и метрическим свойствам. Одним из важных направлений является построение топологических индексов молекулярных структур [1]. Одним из распространенных индексов является число Винера [1-4], аналог которого в теории графов есть дистанция графа. Число Винера используется в задачах установления корреляций типа "структура-свойство", характеризации цикличности и разветвленности молекулярных графов и в других задачах. В [2-4] рассматривается число Винера для классов спиро-конденсированных полициклических структур, для полициклических структур с ациклическими фрагментами и ката-конденсированных структур. Для соединений этих классов сформулированы топологические правила для оценки значений числа Винера при тех или иных структурных перестройках. Топологические правила имеют практическое значение для предсказания упорядочивания изомеров, определения закономерностей между молекулярными свойствами и значениями числа Винера. Одним из недостатков такого подхода является большое количество изучаемых вариантов структур соединений - отсюда и десятки топологических правил. Также при этом требуется знание количества циклов в соединении, размеров циклов и другой дополнительной информации.

В настоящей работе предпринята попытка охарактеризовать изменение дистанции графов при их преобразованиях с более общих позиций, используя методы метрического анализа графов.

1. Основные определения

Пусть G, H – неориентированные связные графы без петель и кратных ребер; $V(G)$ и $V(H)$ – множества вершин G и H; $|V(G)|=p_G$, $|V(H)|=p_H$ – порядки графов. Под расстоянием $d_G(u,v)$ между вершинами $u,v \in V(G)$ понимаем число ребер в кратчайшей простой цепи, соединяющей вершины u и v.

Операцией присоединения графов G и H называется отождествление произвольных вершин $u \in V(G)$, $v \in V(H)$. Граф L, являющийся результатом операции присоединения, обозначим через $L = (G,H,u,v)$ и считаем, что $V(L) = (V(G)\backslash u) \cup (V(H)\backslash v) \cup w$, где вершина $w \in V(L)$ образована при отождествлении вершин u и v так, что $\deg_L(w) = \deg_H(v) + \deg_G(u)$ ($\deg_H(v)$ – степень вершины v в графе H). Обозначим через $D_G(v) = \sum_{u \in V(G)} d_G(u,v)$ дистанцию вершины v в графе G, через $D_G(H) = \sum_{v \in V(H)} D_G(v)$ дистанцию подграфа H в графе G и через $D(G) = \frac{1}{2} \sum_{v \in V(G)} D_G(v)$ дистанцию графа G.

Постановка задачи. Пусть $L = (G,H,u,v)$. Рассмотрим преобразование φ графа L такое, что $\varphi(L) = (G,H,u_1,v_1)$, $u_1 \in V(G)$, $v_1 \in V(H)$ и $L \neq \varphi(L)$. Требуется определить изменение дистанции графа L при осуществлении преобразования φ, т.е. оценить величину $D(L) - D(\varphi(L))$. Содержательно преобразование φ графа L заключается в том, что подграф H "отрывается" от подграфа G и снова присоединяется к нему же, образуя новый граф $\varphi(L)$.

2. Изменение дистанции графа при структурных преобразованиях графа

Пусть $L = (G,H,u,v)$, тогда дистанцию графа L можно определить через дистанции графов G,H и вершин u и v.

ЛЕММА 1. $D(L) = D(G) + D(H) + (p_G-1)D_H(v) + (p_H-1)D_G(u)$.

ДОКАЗАТЕЛЬСТВО. Рассмотрим граф на рис.1. По определению дистанций вершин, подграфов и графов значение $D(L)$ можно представить в следующем виде:

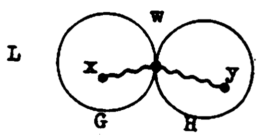

Рис.1

$$D(L) = \frac{1}{2}(D_L(G) + D_L(H) - D_L(w)). \quad (1)$$

Определим значения $D_L(G)$, $D_L(H)$ и $D_L(w)$. Выберем произвольно $x \in V(G)$, тогда можно заметить, что

$$D_L(x) = D_G(x) + \sum_{y\in V(H)\setminus w} (d(x,w) + d(w,y)) = D_G(x) +$$
$$+ (p_H-1)d(x,w) + D_H(v).$$

Суммируя по всем $x\in V(G)$, получаем

$$D_L(G) = 2D(G) + (p_H-1)D_G(u) + p_G D_H(v).$$

Проведя такие же рассуждения для вершин из $V(H)$, имеем

$$D_L(H) = 2D(H) + (p_G-1)D_H(v) + p_H D_G(u).$$

Для вершины $w \in V(L)$ выполняется $D_L(w) = D_G(u) + D_H(v)$. Таким образом, подставляя выражения для $D_L(G)$, $D_L(H)$ и $D_L(w)$ в (I), получаем $D(L) = D(G) + D(H) + (p_G-1)D_H(v) + (p_H-1)D_G(v)$. Лемма доказана.

Из леммы следует, что для определения дистанции графа L достаточно знать значение дистанций вершин в графах G и H.

Обозначим $\varphi(L) = L_1$ и рассмотрим сначала такое φ, что если $L = (G,H,u,v)$, то $L_1 = (G,H,u_1,v)$, т.е. при построении графов L и L_1 в графе H отождествляется одна и та же вершина.

ТЕОРЕМА I. $D(L) - D(L_1) = (p_H-1)(D_G(u) - D_G(u_1))$.

ДОКАЗАТЕЛЬСТВО. По утверждению леммы I дистанции графов L и L_1 можно представить в следующем виде:

$$D(L) = D(G) + D(H) + (p_G-1)D_H(v) + (p_H-1)D_G(u),$$

$$D(L_1) = D(G) + D(H) + (p_G-1)D_H(v) + (p_H-1)D_G(u_1).$$

Отсюда получаем, что $D(L) - D(L_1) = (p_H-1)(D_G(u) - D_G(u_1))$. Теорема доказана.

Если дистанции вершин u и u_1 в графе G неизвестны, но известно расстояние между ними, то изменение дистанции можно оценить через параметры графов G и H и значение расстояния между u и u_1.

СЛЕДСТВИЕ I (локальное перемещение H). Если вершины u и u_1 смежны в G, то $|D(L) - D(L_1)| \le \le (p_G-2)(p_H-1)$.

ДОКАЗАТЕЛЬСТВО. По теореме I, $D(L) - D(L_1) = (p_H-1)(D_G(u) - D_G(u_1))$. Отсюда

$$|D(L) - D(L_1)| = (p_H-1)\,|D_G(u) - D_G(u_1)| =$$
$$= (p_H-1)\Big|\sum_{x\in V(G)\setminus\{u,u_1\}} (d(u,x) - d(u_1,x))\Big| \le$$

$$\leq (p_H-1) \sum_{x\in V(G)\setminus\{u,u_1\}} |d(u,x)-d(u_1,x)| \leq (p_H-1) \sum_{x\in V(G)\setminus\{u,u_1\}} 1 =$$

$$= (p_H-1)(p_G-2).$$

Следствие доказано.

СЛЕДСТВИЕ 2. Если $d_G(u,u_1) = t$, то

$$|D(L) - D(L_1)| \leq (p_H-1)(p_G-t-1)t. \qquad (2)$$

ДОКАЗАТЕЛЬСТВО. Пусть $V_t = \{u, v_1, v_2, \ldots, v_{t-1}, u_1\}$, где v_1, $v_2, \ldots, v_{t-1}$ – вершины кратчайшего пути, соединяющего вершины u и u_1. По теореме I, $D(L)-D(L_1) = (p_H-1)(D_G(u) - D_G(u_1))$. Разобьем дистанцию вершин u и u_1 на два слагаемых следующим образом:

$$D_G(u) = \sum_{v\in V_t} d(u,v) + \sum_{v\in V(G)\setminus V_t} d(u,v),$$

$$D_G(u_1) = \sum_{v\in V_t} d(u_1,v) + \sum_{v\in V(G)\setminus V_t} d(u_1,v).$$

Так как вершины V_t расположены на кратчайшем пути между вершинами u и u_1, то

$$\sum_{v\in V_t} d(u,v) = \sum_{v\in V_t} d(u_1,v) = \frac{1}{2}\, t(t+1).$$

Кроме того, заметим, что для любой $x \in V(G)$ выполняется $|d(u_1,x) - d(u,x)| \leq t$. Отсюда следует, что

$$|D(L)-D(L_1)| = (p_H-1)\,|D_G(u) - D_G(u_1)| =$$

$$= (p_H-1)\,\Big|\sum_{v\in V(G)\setminus V_t} (d(u,v) - d(u_1,v))\Big| \leq$$

$$\leq (p_H-1) \sum_{v\in V(G)\setminus V_t} |d(u,v) - d(u_1,v)| \leq$$

$$\leq (p_H-1) \sum_{v\in V(G)\setminus V_t} t = (p_H-1)t\,|V(G)\setminus V_t| = (p_H-1)t(p_G-t-1).$$

Следствие доказано.

Графы, для которых достигается равенство в (2), приведены на рис.2, $D(L) = 832$, $D(L_1) = 688$.

В ряде задач синтеза молекулярных графов необходимо так осуществлять перестройку структуры графа, чтобы дистанция графа изменялась в заданном интервале, т.е. выполнялось бы неравенство

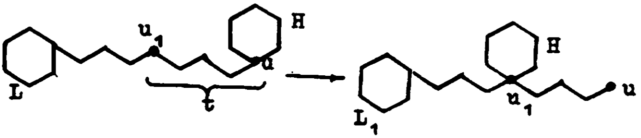

Рис. 2

$|D(L)-D(L_1)| \le b$ для заданного b. В нашей постановке это соот - ветствует тому, что при образовании графа L_1 в графе G могут отождествляться только вершины, удовлетворяющие некоторым условиям. Следующее следствие устанавливает критерий выбора таких вершин.

СЛЕДСТВИЕ 3. Пусть $D_G(u) = a$, тогда для того чтобы выполнялось неравенство $|D(L)-D(L_1)| \le b$, необходимо, чтобы для дистанции вершины u_1 выполнялось

$$a - \frac{b}{p_H-1} \le D_G(u_1) \le a + \frac{b}{p_H-1}.$$

ДОКАЗАТЕЛЬСТВО. По теореме I, $D(L) - D(L_1) = (p_H - 1)(D_G(u) - - D_G(u_1)) = (p_H-1)(a-D_G(u_1))$. Тогда $(p_H-1)|D_G(u_1)-a| \le b$ или

$$-\frac{b}{p_H-1} \le D_G(u_1) - a \le \frac{b}{p_H-1},$$

откуда и следует утверждение следствия.

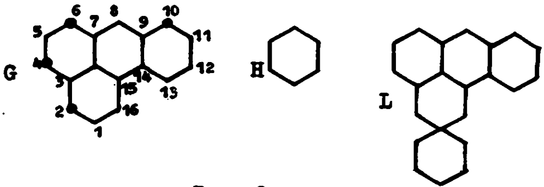

Рис. 3

Рассмотрим для примера граф $L = (G, H, u, v)$ на рис.3 и поставим следующий вопрос: какими способами можно перестроить структуру этого графа так, чтобы $|D(L) - D(L_1)| \le 15$, где $L_1 = (G, H, u_1, v)$? По следствию 3 для вершины-кандидата $u_1 \in V(G)$ должно выполняться неравенство $54 \le D_G(u_1) \le 60$. Этому условию удовлетворяют дистанции вершин графа G, выделенных на рис.3 кружком, так как $D_G(2) = D_G(4) = 58$, $D_G(10) = 56$, $D_G(6) = 54$. Дистанция исходного графа $D(L) = 896$, $D_G(1) = 57$. При $u_1 = 2$ или $u_1 = 4$ дистанция $D(L_1) = 901$, если $u_1 = 6$, то $D(L_1) = 881$, и для $u_1 = 10$ дистанция $D(L_1) = 891$.

ЗАМЕЧАНИЕ. Все вышеприведенные рассуждения остаются в силе и для случая, когда $L_1 = (G, H, u_1, v_1)$, $v_1 \neq v$, если только вершины v и v_1 являются дистанционно эквивалентными в графе H, т.е. если $D_H(v) = D_H(v_1)$.

Рассмотрим преобразование φ такое, что для $L = (G, H, u, v)$ граф $\varphi(L) = L_1 = (G, H, u_1, v_1)$ и $D_H(v) \neq D_H(v_1)$. Представим преобразование графа L в виде двух последовательных преобразований $L \to L^* \to L_1$, где $L^* = (G, H, u_1, v)$. Для каждого из этих преобразований можно использовать предыдущие результаты.

ТЕОРЕМА 2.

$$D(L)-D(L_1)=(p_G-1)(D_H(v)-D_H(v_1)) + (p_H-1)(D_G(u)-D_G(u_1)).$$

ДОКАЗАТЕЛЬСТВО. Для дистанций графов L и L_1 по лемме I имеем $D(L) = D(G)+D(H) + (p_G-1)D_H(v) + (p_H-1)D_G(u)$ и $D(L_1) = D(G) + D(H) + (p_G-1)D_H(v_1) + (p_H-1)D_H(u_1)$. Отсюда получаем выражение для разности дистанций графов L и L_1:

$$D(L) - D(L_1) = (p_G-1)(D_H(v) - D_H(v_1)) + (p_H-1)(D_G(u) - D_G(u_1)).$$

Теорема доказана.

СЛЕДСТВИЕ 4 (о локальном перемещении графов G и H). П у с т ь в е р ш и н ы u и u_1 с м е ж н ы в G, а v и v_1 с м е ж н ы в H. Т о г д а $|D(L)-D(L_1)| \leq 2(p_H-2)(p_G-2) + p_G + p_H - 4$.

ДОКАЗАТЕЛЬСТВО. Представим переход от графа L к графу L_1 как $L \to L^* \to L_1$, где $L^* = (G, H, u_1, v)$. Тогда

$$|D(L)-D(L_1)| = |D(L)-D(L^*)+D(L^*)-D(L_1)| \leq$$
$$\leq |D(L)-D(L^*)| + |D(L^*)-D(L_1)|. \qquad (3)$$

Для оценки разности дистанций используем следствие 2, т.е. $|D(L)-D(L^*)| \leq (p_H-1)(p_G-2)$ и $|D(L^*)-D(L_1)| \leq (p_G-1)(p_H-2)$. Подставляя эти выражения в (3), получаем требуемое неравенство. Следствие доказано.

СЛЕДСТВИЕ 5. Е с л и $d_G(u, u_1) = t$, $d_H(v, v_1) = s$, т о

$$|D(L)-D(L_1)| \leq p_G p_H(t+s)-p_H(t^2+t+s) - p_G(s^2+s+t)+t^2+s^2+t+s. \qquad (4)$$

ДОКАЗАТЕЛЬСТВО. Так же как и ранее, представим преобразование структуры графа L в виде $L \to L^* \to L_1$. Тогда

$$| D(L)-D(L_1) | \le | D(L)-D(L^*) | + | D(L^*)-D(L_1) |. \qquad (5)$$

Для каждой разности дистанций в (5) можно применить следствие 2, т.е. $| D(L)-D(L^*) | \le (p_H-1)(p_G-t-1)t$, $|D(L^*)-D(L_1)| \le (p_G-1)(p_H-s-1)s$. Подставляя это выражение в (5), получаем требуемую оценку. Следствие доказано.

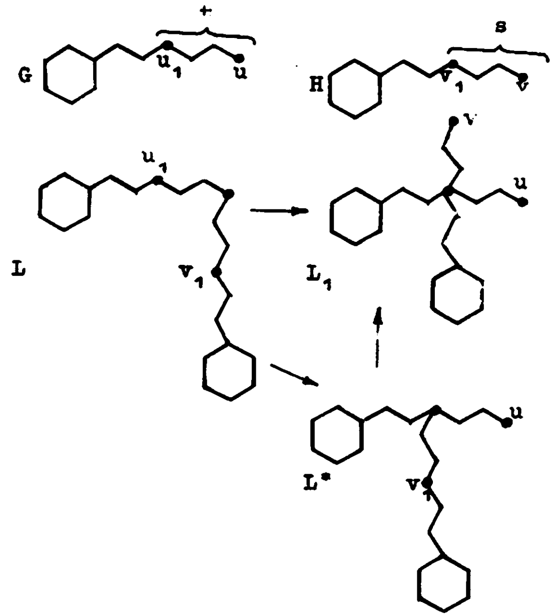

Рис. 4

На рис.4 приведены графы L и L_1, на которых достигается равенство в (4), $D(L) = 1804$, $D(L_1) = 1276$.

Следующее следствие дает возможность осуществлять выбор вершин графов G и H для образования графов с заданным ограничением на изменение дистанции.

СЛЕДСТВИЕ 6. Пусть $D_G(u) = a$, $D_H(v) = c$. Для того чтобы для заданного b выполнялось $|D(L)-D(L_1)| \le b$, дистанции вершин $u_1 \in V(G)$, $v_1 \in V(H)$ должны удовлетворять условию

$$\varphi(a, c, p_G, p_H) - b \le (p_H-1)D_G(u_1)+(p_G-1)D_H(v_1) \le \varphi(a, c, p_G, p_H) + b,$$

г д е $\varphi(a, c, p_G, p_H) = a(p_H-1)+c(p_G-1)$.

ДОКАЗАТЕЛЬСТВО. По теореме 2 имеем $D(L)-D(L_1)=(p_H-1)(D_G(u)-D_G(u_1))+(p_G-1)(D_H(v)-D_H(v_1))$. Отсюда

$$|D(L)-D(L_1)| = |(p_H-1)(a-D_G(u_1)) + (p_G-1)(c-D_H(v_1))| =$$

$$= |a(p_H-1)+c(p_G-1)-(p_H-1)D_G(u_1)-(p_G-1)D_H(v_1)| \le b.$$

Упрощая выражение, получаем условие на дистанции вершин u_1 и v_1

$$a(p_H-1)+c(p_G-1)-b \le (p_G-1)D_H(v_1)+(p_H-1)D_G(u_1) \le a(p_H-1)+c(p_G-1)+b.$$

Следствие доказано.

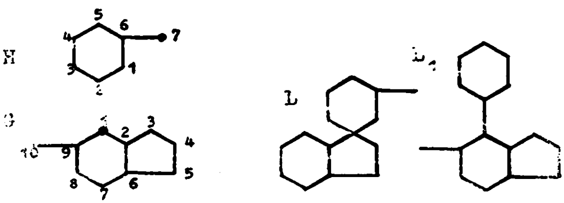

Рис. 5

Рассмотрим в качестве примера граф на рис.5. Какие вершины $u_1 \in V(G)$, $v_1 \in V(H)$ необходимо выбрать, чтобы $|D(L)-D(L_1)| \le 10$? По следствию 6 для таких вершин u_1 и v_1 должно выполняться неравенство $230 \le 9 \cdot D_H(v_1) + 6 \cdot D_G(u_1) \le 250$. Этому условию удовлетворяют, например, дистанции вершин $u_1=1$ и $v_1 = 7$. Дистанция графа $L_1 = (G, H, 1, 7)$ равна $D(L_1) = 399$, а дистанция исходного графа $D(L) = 390$.

3. Алгоритмы построения графов

В этом параграфе будут рассматриваться алгоритмы порождения графов с заданными ограничениями на значение их дистанций. В алгоритме I осуществляется построение множества графов путем присоединения двух заданных графов G и H.

Алгоритм I ($G, H, a, b, GRAPH, SPIRO$)

Вход: I. Графы G, H порядков p_G и p_H.
2. Целые числа a, b.

Выход: Множество графов GRAPH и множество $SPIRO \subseteq V(G) \times V(H)$ такие, что если $L \in GRAPH$, то $L = (G, H, u, v)$, $(u, v) \in SPIRO$ и $a \le D(L) \le b$.

Метод: I. $GRAPH = \emptyset$, $SPIRO = \emptyset$;
2. Для всех вершин $u \in V(G)$, $v \in V(H)$ вычислить значения $D_G(u)$, $D_H(v)$. Пусть $D(G) = n$, $D(H) = m$.
3. Для всех $u \in V(G)$ цикл;
  для всех $v \in V(H)$ цикл;
   I. Вычислить значение $D(L)$ по формуле

   $$D(L) = n+m+D_G(u)(p_H-1) + D_H(v)(p_G-1).$$

   2. Если $a \le D(L) \le b$ то
   $L \to GRAPH$, $(u, v) \to SPIRO$;
  все;
все;

Конец;

Варьируя параметры a и b, алгоритм можно применять для решения различных задач. Например, если необходимо построить все графы, образованные из G и H операцией присоединения, и вычислить их дистанции, то достаточно положить $a = 0$ и $b = (p_G+p_H)^3$. Если нужно определить, существует ли граф L, образованный из G и H, с дистанцией $D(L) = k$, то полагаем $a = b = k$. Для того чтобы множество GRAPH содержало только неизоморфные графы, на шаге 3 алгоритма I должны использоваться только вершины-представители орбит группы автоморфизмов графов G и H. Дистанционно эквивалентные вершины можно отбрасывать на шаге 2. Однако одинаковым значениям дистанций могут соответствовать различные по структуре графы, и если одни разновидности структуры предпочтительнее других, то необходимо порождать все графы для последующего анализа. На рис.6 показаны неизоморфные графы, построенные по алгоритму I при $a=0$, $b = (p_G+p_H)^3$ с дистанциями $D(L) = 1832$, $D(L_1) = 1752$, $D(L_2) = 1632$, $D(L_3) = 1592$, $SPIRO = \{(1,1),(2,1),(3,1)\}$.

В алгоритме 2 осуществляется построение графов, значения дистанции которых отличаются от значения дистанции заданного графа в указанном интервале.

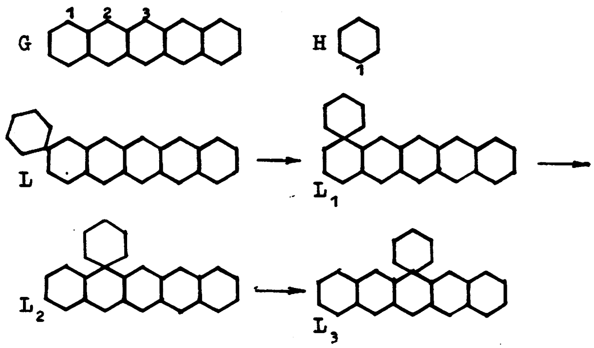

Рис. 6

Алгоритм 2 (G, H, L, a, b, GRAPH, SPIRO)

Вход: I. Графы G, H порядков p_G и p_H.

2. Граф $L = (G, H, u, v)$, $D(L) = k$.

3. Целые числа a, b.

Выход: Множество графов GRAPH и множество SPIRO $\subseteq V(G) \times V(H)$ такие, что если $L_1 \in$ GRAPH, то $L_1 = (G, H, u_1, v_1), (u_1, v_1) \in$ SPIRO и $a \le D(L) - D(L_1) \le b$.

Метод: I. GRAPH $= \emptyset$, SPIRO $= \emptyset$;

2. Алгоритм I ($G, H, k-b, k-a$, GRAPH, SPIRO);

3. GRAPH $\leftarrow$ GRAPH$\setminus L$, SPIRO $\leftarrow$ SPIRO$\setminus(u, v)$;

Конец;

В зависимости от значений параметров a и b можно получать графы с дистанцией, меньшей или большей, чем $D(L)$. Если положить $a < 0$, $b = 0$, то для любого графа $L_1 \in$ GRAPH $D(L_1) \ge D(L)$, если же $a = 0$, $b > 0$, то выполняется $D(L_1) \le D(L)$. Рассмотрим граф L на рис.6, и пусть $a = 130$, $b = 220$, тогда результатом работы алгоритма 2 будет граф L_2 на рис.6, SPIRO $= \{(2, 1)\}$.

Заключение

В настоящей работе рассматривалось изменение дистанции при преобразованиях структуры графа, образованного из пары заданных графов G и H отождествлением одной вершины. Преобразование состоит в том, что в графах G и H отождествляются новые вершины. На структуру G и H не накладывается никаких ограничений, вследствие чего формулы, описывающие изменение дистанции графа, носят достаточно общий характер. Явно определяя структуру графов G и H, можно конкретизировать вид формул. Полученные результаты служат основой для алгоритмического подхода к определению дистанционных изменений по матрице слоев (матрице расстояний, матрице смежности) графа.

Первый из авторов выражает благодарность профессору Поланскому О.Е. и профессору Бончеву Д. за предоставленную возможность ознакомиться с материалами по теории числа Винера.

Авторы выражают особую благодарность профессору Бончеву Д., общение с которым стимулировало их интерес к данной проблеме.

Литература

I. ROUVRAY D.H. Should we have designs on topological indices? - Chemical applications of topology and graph theory, 1983, v.28, p.159-177.

2. MEKENYAN O., BONCHEV D., TRINAJSTIĆ N. Topological rules for spirocompounds.- Math.chem.(MATCH), 1979, N 6, p.93-115.

3. MEKENYAN O., BONCHEV D., TRINAJSTIĆ N. A topological characterization of cyclic structures with acyclic branches.- Math. Chem. (MATCH), 1981, N 11, p.145-168.

4. MEKENYAN O., BONCHEV D. Structural Complexity and Molecular Properties of Cyclic Systems with Acyclic Branches.- Croat. Chem.Acta, 1983, v.56, p.237-261.

Поступила в ред.-изд.отд.
23 июля I986 года