МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ХИМИЧЕСКОЙ ИНФОРМАТИКЕ
(Вычислительные системы)

1991 год

Выпуск 140

УДК 519.1

# МЕТРИЧЕСКИЕ ИНВАРИАНТЫ ПОДГРАФОВ МОЛЕКУЛЯРНЫХ ГРАФОВ

А.А.Добрынин, В.А.Скоробогатов

## Введение

Метрические инварианты графов,определяемые как функции от расстояний между вершинами графа,находят многочисленные приложения в областях,для которых изучаемые объекты или отношения между ними моделируются графами или их обобщениями. К таким областям можно отнести социологию, химию, биологию, экологию и др. [1-3]. Метрические инварианты, именуемые топологическими индексами, используются для установления подобия и характеризации структур молекулярных графов в органической химии. Известны десятки топологических индексов, значения которых на тех или иных классах молекулярных графов дают хорошие корреляции со свойствами соответствующих химических соединений. Результаты подобных исследований используются для прогнозирования свойств и конструирования соединений с заданными свойствами,например, новых лекарственных препаратов и т.п.Обзорные материалы по топологическим индексам и их применениям можно найти в [3-15].

Исторически сложилось, что топологические индексы,как правило, являются интегральными метрическими инвариантами молекулярных графов. В то же время при решении практических задач возникает необходимость анализа молекулярного графа не только как целого, но и характеризации его отдельных частей. При изучении взаимозависимости между структурой и активностью химических соединений степень проявления активности может зависеть как от расположения определенных фрагментов в молекуле, так и от их взаимного расположения и характеристик всей молекулы. Так, взаимодействие лекарственного вещества с организмом на разных стадиях определяется интегральными физико-химическими свойствами молекул препарата (транспортировка молекул, проникновение через мембраны и т.п.) и локальными характеристиками фармакофоров - подструктур, содержащих функциональные группы, расположение которых в молекуле должно отвечать некоторым требованиям для обеспечения специфического биологического действия. Например, синтезированные производные халконов, изображенные на рис. 1 в виде молекулярных графов на плоскости, отличаются друг

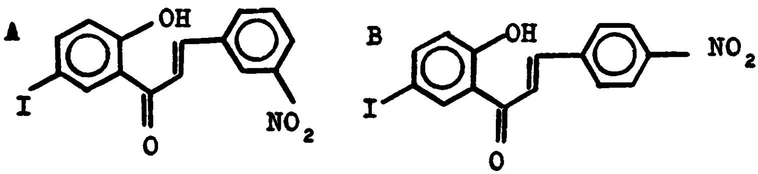


Рис. 1

от друга изменением присоединения группы $NO_2$ к ароматическому кольцу. Токсичность и способность воздействовать на микроорганизмы для соединения **A** резко усиливаются по сравнению с соединением **B**. Отличие в свойствах соединений может быть обусловлено изменением положения функциональной группы $NO_2$ относительно всех вершин молекулярного графа, относительно какого-либо другого фрагмента графа или нескольких фрагментов, отвечающих химическим функциональным группам.

Один из первых подходов к определению топологических индексов фрагментов и схем их вычисления был предложен в [16, 17]. В [17] введены новые топологические индексы для описания фрагментов молекулярных графов, модифицирован целый ряд известных топологических индексов для этих целей и изучено их поведение для некоторых семейств молекулярных структур. Настоящая работа посвящена развитию подхода из [16]. Будут рассмотрены метрическая характеризация положения подграфа в графе, взаимного положения подграфов и положения подграфа в семействе графов, обсуждены особенности вычислительных схем соответствующих метрических инвариантов. Изложение ориентировано на химические приложения, поэтому, наряду с теоретико-графовыми терминами "граф" и "подграф", будут использоваться их синонимы - молекулярный граф и фрагмент молекулярного графа.

## §1. Типы подграфов и метрических инвариантов

Рассматриваются связные непомеченные графы $G(V,E)$ без петель и кратных ребер с множеством вершин $V(G)$ и множеством ребер $E(G)$. Число вершин $p(G)$ графа называется его *порядком*, а число ребер $q(G)$ - *размером* графа. Химические соединения органической химии моделируются молекулярными графами, помеченные вершины которых соответствуют атомам соединения, а веса (кратности) ребер описывают тип химической связи между атомами. Далее в основном будут использоваться непомеченные скелеты молекулярных графов. В каждом случае, когда наличие меток вершин и/или кратностей ребер является существенным, тип графа оговаривается. Все термины и понятия, явно не определяемые в тексте, определены в [18].

В графах будем выделять два основных типа подграфов. Подграф $H(V',E')$ графа $G$ называется *порожденным* подмножеством вершин $V' \subseteq V(G)$, если число ребер в подмножестве $E' \subseteq E(G)$ является максимальным, т.е. любое ребро $(v,u) \in E(G)$ для $v,u \in V'$ принадлежит и множеству $E'$. Порожденный подграф на множестве $V'$ обозначим $\langle V'\rangle$. Под *частичным* подграфом $H(V',E')$ понимается подграф с множеством вершин $V' \subseteq V(G)$ и произвольным множеством ребер $E \subset E(\langle V'\rangle)$. Частичный подграф можно также задать как подграф, порожденный подмножеством ребер $E' \subset E(G)$, множество вершин которого образуют все инцидентные ребрам из $E'$ вершины. Заметим, что последнее определение не допускает появления в частичном подграфе изолированных вершин. На произвольном подмножестве вершин графа существуют единственный порожденный подграф и целое семейство частичных подграфов, отличных от порожденного (считаем,что в случае вполне несвязного порожденного подграфа частичных подграфов на этом множестве вершин не существует). Подграфы связного графа могут быть как связные, так и несвязные. Рассматриваемый как граф, несвязный подграф состоит из нескольких компонент связности. На рис. 2 приводятся примеры связных и несвязных порожденных и частичных подграфов одного и того же графа. Основным объектом дальнейшего рассмотрения являются порожденные несвязные подграфы.

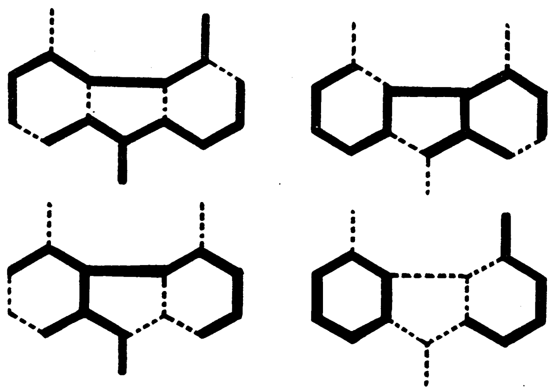


Рис. 2

Под расстоянием в графе будем понимать меру близости между вершинами графа, не обязательно удовлетворяющую аксиомам метрики. Рассмотрим три типа расстояний. *Естественным расстоянием* $d_G(v,u)$ между вершинами $v,u \in V(G)$ в графе $G$ называется длина кратчайшей по числу ребер простой цепи, соединяющей $v$ и $u$ в $G$. *Протяженность* $el_G(v,u)$ между вершинами $v,u \in V(G)$ определим как длину длиннейшей по числу ребер простой цепи между вершинами $v$ и $u$ в графе $G$. *Цепное расстояние* $p_G(v,u)$ между вершинами $v$, $u \in V(G)$ есть сумма длин всех простых цепей, соединяющих $v$ и $u$ в графе $G$. Естественное расстояние и протяженность удовлетворяют аксиомам метрики [18,19], в то время как цепное расстояние не является метрикой из-за нарушения неравенства треугольника.

Введенные расстояния порождают соответствующие локальные и интегральные метрические инварианты графов. Инварианты, построенные с использованием цепного расстояния и протяженности, будем называть цепными инвариантами [20]. Приведем перечень метрических и цепных инвариантов, которые будут применяться для описания положения подграфов в графе. Рассматриваемые инварианты можно подразбить на два различных по свойствам класса - эксцентриситетные и дистанционные инварианты, построенные на основе дистанций и эксцентриситетов вершин графа [16,19-22].

Вершинные (локальные) метрические и цепные инварианты:

1) *дистанция вершины* в графе определяется как

$$D(v) = \sum_{u \in V(G)} d_G(v,u);$$

2) *эксцентриситет вершины* $v$ в графе $G$ есть величина

$$e(v) = \max_{u \in V(G)} d_G(v,u);$$

3) *цепная дистанция вершины* $v$ в графе $G$ задается формулой

$$D_\tau(v) = \sum_{u \in V(G)} p_G(v,u);$$

4) *цепной эксцентриситет вершины* $v$ (число протяженности)

$$e_\tau(v) = \max_{u \in V(G)} el_G(v,u);.$$

Построенные с использованием вершинных инвариантов интегральные метрические и цепные инварианты:

1) *дистанция графа* $G$ есть сумма дистанций вершин

$$D(G) = \sum_{v \in V(G)} D_G(v);$$

2) *э сцентриситетом графа* $G$ называется величина

$$e(G) = \sum_{v \in V(G)} e_G(v);$$

3) *цепная дистанция графа* $G$ задается формулой

$$D_\tau(G) = \sum_{v \in V(G)} D_\tau(v);$$

4) *цепной эксцентриситет графа* $G$ определяется выражением

$$e_\tau(G) = \sum_{v \in V(G)} e_\tau(v).$$

В классе деревьев цепное расстояние совпадает с естественным и является метрикой, и по построению цепных инвариантов их значения совпадают со значениями метрических инвариантов. Цепные инварианты более полно отражают особенности структуры молекулярных графов, так как учитывают степень разветвленности и цикличности графа. Поэтому, наряду с метрическими, цепные инварианты успешно используются для установления подобия структур молекулярных графов и прогнозирования свойств соединений [4,23-25]. В настоящей работе выражение для дистанции графа

$D(G)$ отличается от аналогичного в [3-16] множителем, равным 2, поэтому для известного топологического индекса - числа Винера $W(G)$ - выполняется $2W(G) = D(G)$. Число Винера $W(G)$ является, вероятно, наиболее изученным топологическим индексом, для которого установлены корреляции между его значениями и свойствами химических соединений, например, полициклических ароматических структур [26-30]. В [31-39] исследовались математические свойства индекса Винера для классов молекулярных графов, поведение индекса при структурных преобразованиях графов. Рассматривались также различные модификации и обобщения индекса Винера и возможности их использования для предсказания свойств соединений [40].

Укажем еще несколько используемых далее метрических инвариантов: радиус и диаметр графа есть $r(G) = \min\limits_{v \in V(G)} e(v)$ и $d(G) = \max\limits_{v \in V(G)} e(v)$ соответственно. *Периферией графа* мощности $n$ для некоторого инварианта назовем подграф на $n$ вершинах, на котором инвариант достигает максимального или "близкого" к нему значения. *Центром графа* мощности $n$ для инварианта назовем подграф на $n$ вершинах, на котором инвариант достигает минимального или "близкого" к нему значения.

## §2. Требования к топологическим индексам фрагментов

К общим требованиям для построения метрических инвариантов подграфов естественно отнести необходимость характеризации инвариантами: а) топологии фрагмента $F$; б) топологии графа $G$; в) связей фрагмента $F$ с оставшейся частью $G-F$ графа.

Ниже дается более детальная возможная интерпретация перечисленных требований к инвариантам подграфов.

1. <u>Монотонность изменения значений инварианта $f$ при увеличении порядка подграфа</u>. Если инвариант является возрастающей

функцией, то для подграфов $F_1 \subset F_2 \subseteq G$ выполняется неравенство $f(F_1) < f(F_2)$. Отсюда следует, что для произвольного подграфа $F \subset G$ значение его метрического инварианта всегда меньше значения инварианта для всего графа - $f(F) < f(G)$.

2. Монотонность изменения значений метрического инварианта в зависимости от положения фрагмента в графе. Пусть фрагмент $F$ расположен в графе $G$ несколькими способами $F_1$, $F_2, \ldots, F_n$ так, что $F_i$ находится "ближе" к центру графа $G$, чем $F_{i-1}$ для $i = 1,2,\ldots,n$ (рис.3), тогда выполняются неравенства $f(F_1) \leq f(F_2) \leq \ldots \leq f(F_n)$. Для приложений достаточно потребовать, чтобы "существенно" различные положения фрагмента в графе не имели бы близких значений инварианта.

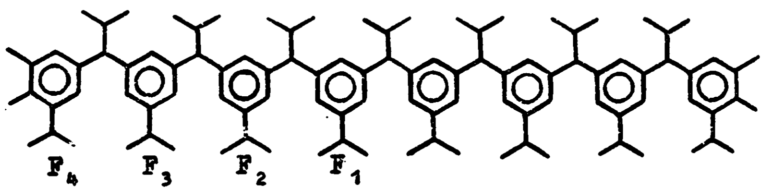



Рис. 3

3. Требование выполнения условия $f(F_1) \leq f(F_2)$ для произвольных фрагментов $F_1$ и $F_2$ порядков $p(F_1) \leq p(F_2)$ не является естественным, так как фрагмент меньшего порядка $F_1$, расположенный на периферии графа, может иметь большее значение индекса, чем фрагмент $F_2$ большего порядка, но находящийся в центре графа.

4. Метрическая характеризация связи фрагмента с остальной частью графа. Абсолютные значения инвариантов фрагмента $F$ могут нести недостаточно информации о том, какой является оставшаяся часть $G-F$ графа. Поэтому необходимо учитывать не

только значение инварианта фрагмента $f(F)$, но и значения $f(G-F)$, $f(G)-f(F)$ и т.п.

5. Нормировка значений метрического инварианта фрагмента. Естественной нормировкой является отношение $f(F)$ к значению инварианта для всего графа. В этом случае для подграфа $F \subseteq G$ выполняется $0 \leq f(F)/f(G) \leq 1$. Нулевое значение в неравенстве может достигаться на подграфе $F \subset G$, являющимся изолированной вершиной в $G$, правая граница достигается при совпадении $F$ и $G$. При такой нормировке также учитывается значение инварианта для графа в целом. Далее будут использоваться и другие виды нормировки.

6. Не очень высокая вырождаемость метрического инварианта для фрагментов. Инвариант должен быть достаточно чувствительным к неэквивалентным положениям фрагмента в графе. Например, такой простой инвариант, как число ребер фрагмента вырождается на всех фрагментах с одинаковым числом ребер вне зависимости от места их расположения в графе.

7. По возможности единообразная схема характеризации подграфов различных типов - связных и несвязных, порожденных и частичных и других.

8. При описании положения подграфа в графе относительно другого подграфа значения метрических инвариантов должны учитывать возможность пересечения подграфов по вершинам и ребрам.

9. Для возможности сравнения положения подграфа в нескольких графах различия в параметрах графов (разное число вершин, ребер, циклов и т.п.) не должны оказывать влияние на значения метрических инвариантов фрагмента.

В зависимости от характера решаемых задач перечисленные требования к инвариантам фрагментов могут быть дополнены или заменены другими.

## §3. Положение подграфа в графе

В данном параграфе рассматривается возможность использо - вания метрических инвариантов для приближенного описания включения фрагментов в структуру графов, позволяющего оценивать положение подграфа в графе. Определим метрические и цепные инварианты для подграфов.

*Дистанция подграфа* $F \subseteq G$ в графе $G$ есть величина

$$D_G(F) = \sum_{\substack{v \in V(F) \\ u \in V(G)}} d_G(u,v).$$

*Эксцентриситет подграфа* $F \subseteq G$ в графе $G$ определяется как

$$e_G(F) = \sum_{v \in V(F)} \max_{u \in V(G)} d_G(u,v).$$

*Нормированная дистанция подграфа* $F \subseteq G$ задается выражением

$$D(F,G) = \frac{1}{D(G)} D_G(F).$$

*Нормированный эксцентриситет подграфа* $F \subseteq G$ есть функция

$$e(F,G) = \frac{1}{e(G)} e_G(F).$$

Цепные инварианты $D_{\tau_G}(F)$, $e_{\tau_G}(F)$, $D_\tau(F,G)$ и $e_\tau(F,G)$ фрагмента $F$ определяются аналогично метрическим.

Такого вида метрические инварианты, называемые в [16] от - носительными метрическими характеристиками, позволяют описать подграф, учитывая характеристики графа в целом.

В метрических и цепных инвариантах учитываются связи вершин подграфа со всеми вершинами графа. Поэтому можно считать, что чем ближе для подграфа $F$ значения функций $e_G(F)$ и $D_G(F)$, $e_{\tau_G}(F)$ и $D_{\tau_G}(F)$ к минимальному значению среди всех подграфов того же порядка, что и $F$, тем более "центральное" положение подграф $F$ занимает в графе, а если значения инвариантов близки к максимальному значению, то $F$ расположен "ближе" к периферии графа. В табл.1 приводятся результаты вычислений метрических и цепных инвариантов для фрагментов некоторых молекулярных графов. Для подграфов $F_1$ и $F_4$, $F_2$ и $F_3$ в графе $G$ на рис.4 значения метрических

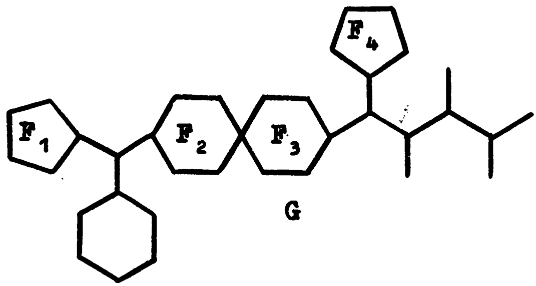
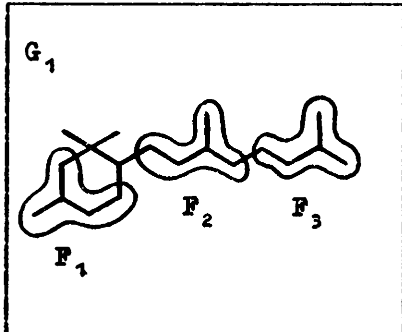


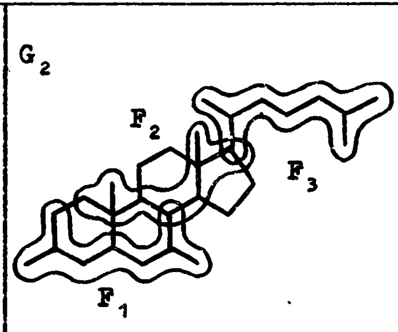

Рис. 4

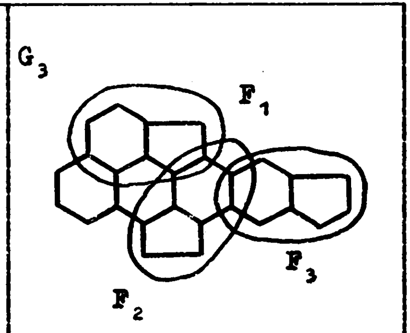
инвариантов совпадают: $D(F_1,G) = D(F_4,G) = 0.1593$ и $D(F_2,G) = D(F_3,G) = 0.1311$, $e(F_1,G) = e(F_4,G) = 0.1592$ и $e(F_2,G) = e(F_3,G) = 0.1278$, а значения цепных инвариантов различаются: дистанции $D_\tau(F_1,G) = 0.1822$, $D_\tau(F_4,G) = 0.2234$, $D_\tau(F_2,G) = 0.0924$ и $D_\tau(F_3,G) = 0.1055$, цепные эксцентриситеты $e_\tau(F_1,G) = 0.1525$, $e_\tau(F_4,G) = 0.1615$, $e_\tau(F_2,G) = 0.1379$ и $e_\tau(F_3,G) = 0.1470$.

Таблица 1

Метрические инварианты подграфов молекулярных графов




| Функции | Граф $G_1$ | | | Граф $G_2$ | | | Граф $G_3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $F_1$ | $F_2$ | $F_3$ | $F_1$ | $F_2$ | $F_3$ | $F_1$ | $F_2$ | $F_3$ |
| $D(F,G)$ | 0.2717 | 0.2151 | 0.3235 | 0.3174 | 0.2574 | 0.3665 | 0.3069 | 0.2810 | 0.3634 |
| $D_\tau(F,G)$ | 0.3080 | 0.1883 | 0.2861 | 0.3472 | 0.2614 | 0.3164 | 0.3102 | 0.2779 | 0.3694 |
| $e(F,G)$ | 0.2905 | 0.2011 | 0.3017 | 0.3354 | 0.2733 | 0.3416 | 0.3099 | 0.2723 | 0.3662 |
| $e_\tau(F,G)$ | 0.2765 | 0.2028 | 0.2949 | 0.3124 | 0.3124 | 0.3008 | 0.3333 | 0.3333 | 0.3333 |

## §4. Положение одновершинного фрагмента

Изучение метрических инвариантов одновершинного фрагмента графа представляет особый интерес, так как одновершинный фрагмент является наименьшим фрагментом в графе. Метрические инварианты подграфа получаются суммированием соответствующих метрических характеристик его вершин, следовательно, по значениям инвариантов одновершинного фрагмента можно судить о возможных значениях метрических инвариантов для подграфа произвольного порядка. Кроме того, на примере одновершинного фрагмента удобно изучать поведение тех или иных метрических инвариантов в зависимости от положения подграфа в графе.

По построению рассматриваемые нами метрические инварианты подграфов зависят как от дистанции и эксцентриситетов вершин, так и от дистанции и эксцентриситета графа. Укажем некоторые известные оценки значений этих характеристик:

1) $2p(p-1)-2q \le D(G) \le \frac{1}{3}(p^3+5p-6)-2q$ [41];

2) $2(p-1) \le D(G) \le (p-1)(p+2)-2q$ [41];

3) $2(2p-\deg(v)-2) \le D(v) \le (p-1)(p+2)-2q$ [41];

4) $$D(G) \ge \begin{cases} \frac{1}{3}d(d+1)(d+2)+\frac{1}{2}(p-d-1)(2p+d^2+1) & \text{при } d \text{ нечетном}, \\ \frac{1}{3}d(d+1)(d+2)+\frac{1}{2}(p-d-1)(2p+d^2) & \text{при } d \text{ четном}, \end{cases}$$

где $d$ есть диаметр графа [42];

5) $D(G) \le \frac{1}{3}p(p-1)(p+1)$ [43];

6) если $G$ является двусвязным графом (т.е. графом без точек сочленения), то $D(G) \le 2p\lfloor \frac{1}{4}p^2 \rfloor$, где $\lfloor p \rfloor$ означает наибольшее целое, не превосходящее $p$ [42];.

7) $1 \le r(G) \le e(v) \le d(G) \le p-1$;

8) $e(v) \le \lfloor \frac{1}{3}(2p-2) \rfloor$ [43];

9) $p \le e(G) \le p^2-2q$ [44].

Для каждого из указанных выше неравенств существуют графы, на которых достигаются их верхние и нижние границы.

Приведенные величины можно использовать для оценки значений метрических инвариантов $e(v)/e(G)$ и $D(v)/D(G)$, однако для более точных оценок следует учитывать взаимозависимость значений $e(v)$ и $e(G)$, $D(v)$ и $D(G)$ в графе.

Оценки очевидны в случае равенства значений метрических инвариантов для всех вершин в графе (графы с транзитивной группой автоморфизмов и др.), т.е. если для произвольных вершин $v,u \in V(G)$ выполняется $D(v) = D(u)$ или $e(v) = e(u)$, то $D(v)/D(G) = e(v)/e(G) = 1/p$. Следующие неравенства являются уточнением оценок отношения $e(v)/e(G)$ из [16].

УТВЕРЖДЕНИЕ 1. *Для вершины* $v \in V(G)$ *в графе* $G$ *выполняется*

$$\frac{1}{2p-1} \le \frac{e(v)}{e(G)} \le \frac{2}{p+2} .$$

ДОКАЗАТЕЛЬСТВО. Для суммы эксцентриситетов вершин можно записать

$$e(G) = \sum_{v \in V(G)} e(v) \ge 2d(G)+r(G)(p-2)$$

в силу выполнения $r(G) \le d(G)$. Далее имеем

$$\frac{e(G)}{e(v)} \ge \frac{2d(G)+r(G)(p-2)}{d(G)} = 2+ \frac{r(G)}{d(G)}(p-2) \ge 2+ \frac{1}{2}(p-2),$$

так как $d(G) \le 2r(G)$ . Отсюда следует верхняя оценка. Для нижней оценки аналогично имеем

$$\frac{e(G)}{e(v)} \le \frac{r(G)+d(G)(p-1)}{r(G)} = 1+\frac{d(G)}{r(G)}(p-1) \le$$
$$\le 2(p-1)+1 = 2p-1 .$$

Утверждение доказано.

На рис.5 приводятся графы, на которых достигаются верхняя и нижняя оценки утверждения. Действительно, из доказательства следует, что верхняя оценка достигается на графах, в которых две вершины являются диаметральными, а остальные - радиальными и выполняется $2r(G) = d(G)$ . Таким свойством обладают, например, полные графы порядка $p$ с удаленным ребром $K_p - x$. Нижняя оценка достигается на графах с единственной радиальной (центральной) вершиной и остальными диаметральными вершинами и $2r(G) = d(G)$ . Звезда $K_{1,p-1}$ и колесо $W_p$ удовлетворяют этим свойствам, как и другие подобные им графы на рис.5. В табл. 2 приводятся численные значения метрического инварианта $e(v)/e(G)$ в графах, представляющих молекулярные структуры химических соединений.

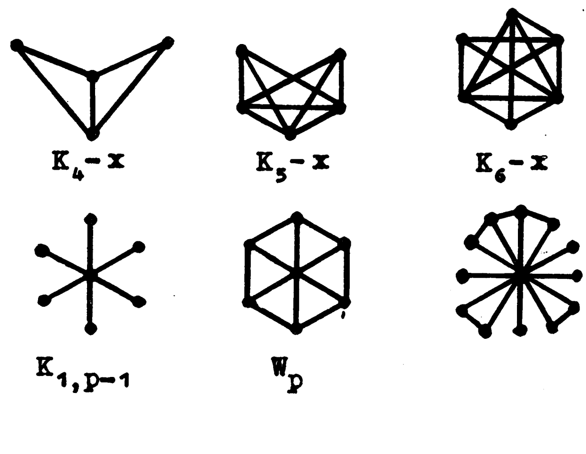
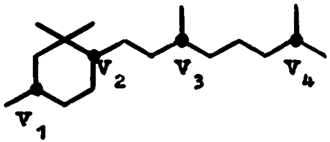



Рис. 5

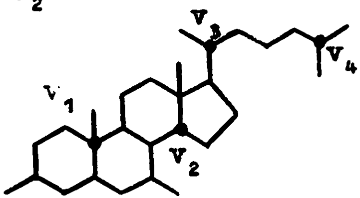
Т а б л и ц а 2

Метрические инварианты одновершинных подграфов молекулярных графов

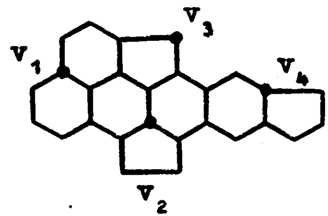
$G_1$




$G_2$




$G_3$




| Функции | Граф $G_1$ | | | | Граф $G_2$ | | | | Граф $G_4$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_1$ | $v_2$ | $v_3$ | $v_4$ |
| $D(v,G)$ | .0566 | .0392 | .0410 | .0651 | .0302 | .0257 | .0307 | .0495 | .0388 | .0290 | .0330 | .0405 |
| $D_\tau(v,G)$ | .0584 | .0294 | .0371 | .0579 | .0336 | .0248 | .0294 | .0407 | .0321 | .0345 | .0378 | .0381 |
| $e(v,G)$ | .0615 | .0447 | .0391 | .0615 | .0354 | .0257 | .0322 | .0450 | .0423 | .0282 | .0282 | .0423 |
| $e_\tau(v,G)$ | .0507 | .0369 | .0415 | .0599 | .0377 | .0343 | .0309 | .0377 | .0370 | .0370 | .0370 | .0370 |

Рассмотрим дистанционные характеристики одновершинного фрагмента в графе.

УТВЕРЖДЕНИЕ 2. *Для вершины* $v \in V(G)$ *графа* $G$ *порядка* $p$ *выполняются неравенства*

$$\frac{1}{2p} \le \frac{D(v)}{D(G)} \le \sqrt{2}\,\frac{1}{\sqrt{p}} .$$

ДОКАЗАТЕЛЬСТВО. Нижняя оценка сразу следует из неравенств $D(G) = \sum\limits_{u,v \in V(G)} d(u,v) \le \sum\limits_{u,v \in V(G)} (d(u,w) + d(w,v)) \le$ $\le 2pD(w)$ для некоторой вершины $w$. Так как выбор вершины $w$ произволен, то положим $D(w) = \min\{D(v) \mid v \in V(G)\}$. Тогда для любой вершины $v \in V(G)$ выполняется $\frac{D(v)}{D(G)} \ge$ $\ge \frac{D(w)}{2pD(w)} = \frac{1}{2p}$. Для получения верхней оценки запишем

$$\frac{D(v)}{D(G)} \le \frac{\max\{D(u) \mid u \in V(G)\}}{p \cdot \min\{D(u) \mid u \in V(G)\}}$$

и далее оценим отношение наименьшей и наибольшей дистанций вершин в графе.

Пусть $Q$ есть множество графов порядка $p$ и диаметра $d$. Обозначим $D_{min} = \min\limits_{G \in Q} \min\limits_{v \in V(G)} D(v)$ и $D_{max} =$ $= \max\limits_{G \in Q} \max\limits_{v \in V(G)} D(v)$ .

ЛЕММА. *Для определенных выше величин выполняются равенства* $D_{max} = pd - \frac{1}{2}d(d+1)$ *и* $D_{min} = \frac{1}{2}d(\frac{1}{2}d-1) + p - C$ *, где* $C = 1$ *, если* $d$ *четно, и* $C = 3/4$*, если* $d$ *нечетно.*

ДОКАЗАТЕЛЬСТВО. В графе диаметра $d$ существует кратчайшая цепь длины $d$ между диаметральными вершинами. Среди вершин простой цепи наименьшую дистанцию имеет центральная вершина $v$. Пусть оставшиеся $p-d-1$ вершин, не входящих в простую цепь длины $d$, располагаются на наименьшем расстоянии от вершины $v$. Из построения следует, что $D(v) = D_{min}$ и $D(v) = \frac{1}{2}\cdot\frac{d}{2}\left(\frac{d}{2}+1\right)+\frac{1}{2}\cdot\frac{d}{2}\left(\frac{d}{2}+1\right)+(p-d-1)$, если $d$ четно, и $D(v)=\frac{1}{2}\frac{d-1}{2}\left(\frac{d-1}{2}+1\right)+\frac{1}{2}\frac{d-1}{2}\left(\frac{d-1}{2}+1\right)+$ $+(p-d-1)$, если $d$ нечетно. После приведения выражений получим требуемые значения. Наибольшую дистанцию среди вершин простой цепи имеет концевая вершина $v$. Пусть остальные $p-d-1$ вершины, не принадлежащие цепи, расположены на расстоянии $d$ от вершины $v$. Для такой вершины $D(v) = D_{max}$ и

$$D(v)=\frac{1}{2}d(d+1)+d(p-d-1)=pd-\frac{1}{2}d(d+1).$$

Для завершения доказательства покажем, что такие графы существуют. Для вершин графов на рис.6 выполняются $D(v) = D_{min}$ и $D(u) = D_{max}$. Лемма доказана.

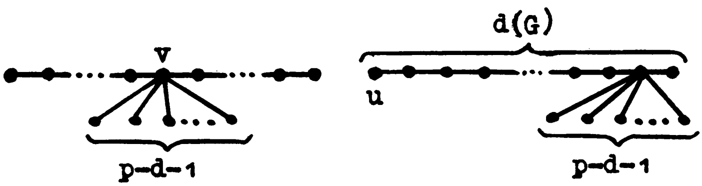



Рис. 6

Оценим по лемме значение отношения $D_{min}/D_{max}$. Имеем

$$\frac{D_{min}}{D_{max}}=\frac{d^2/4-d/2+p-C}{pd-d^2/4-d/2}\geq\frac{d^2/4+p/2}{pd}=$$

$$=\frac{1}{p}\left(\frac{d}{4}+\frac{p}{2d}\right)\geq\frac{1}{2p}\cdot\min_{d}\left(\frac{d}{2}+\frac{p}{d}\right)=\frac{1}{p}\left(\frac{\sqrt{2p}}{2}+\frac{p}{\sqrt{2p}}\right)=\frac{1}{\sqrt{2}}\cdot\frac{1}{\sqrt{p}}.$$

Отсюда следует верхняя оценка вследствие выполнения неравенства

$$\frac{D(v)}{D(G)} \le \frac{\max\{D(u) \mid u \in V(G)\}}{p \cdot \min\{D(u) \mid u \in V(G)\}} \le$$

$$\le \frac{D_{max}}{pD_{min}} \le \frac{\sqrt{2}\sqrt{p}}{p} = \sqrt{2}\frac{1}{\sqrt{p}}.$$

Утверждение доказано.

Из доказательства утверждения следует неравенство для экстремальных значений дистанций вершин в графах - максимальная дистанция вершин в графе всегда меньше удвоенной минимальной дистанции $D_{max} < 2D_{min}$ . Приведем примеры графов, в которых для одновершинных фрагментов значения метрических характеристик близки к верхним и нижним границам утверждения. Так как для звезды $D(K_{1,p-1}) = 2(p-1)^2$, то для центральной вершины $v$ звезды имеем

$$\frac{D(v)}{D(K_{1,p-1})} = \frac{p-1}{2(p-1)^2} = \frac{1}{2(p-1)},$$

а для центральной вершины колеса $W_p$ получим

$$\frac{D(v)}{D(W_p)} = \frac{p-1}{2(p-1)(p-2)} = \frac{1}{2(p-2)}.$$

Для графа с диаметром $d = \sqrt{p}$ на рис. 6 справа выполняется $\frac{D(u)}{D(G)} \ge C\frac{1}{\sqrt{p}}$ при $p > 2$ и константе $C < 1$ ,что устанавливается непосредственным вычислением. Для концевой вершины $v$ простой цепи $\frac{D(v)}{D(G)} = \frac{3}{2(p+1)}$ , а для центральной вершины $u$ цепи с нечетным числом вершин $\frac{D(u)}{D(G)} = \frac{3}{4p}$.

В табл. 2 приводятся численные результаты значений дистанционных метрических характеристик одновершинных фрагментов в молекулярных графах.

Проиллюстрируем взаимное влияние метрических инвариантов $D(v)$ и $D(G)$ в графе. Для этого рассмотрим семейство графов $G_n$ на рис.7, состоящих из простой цепи длины $n$, к

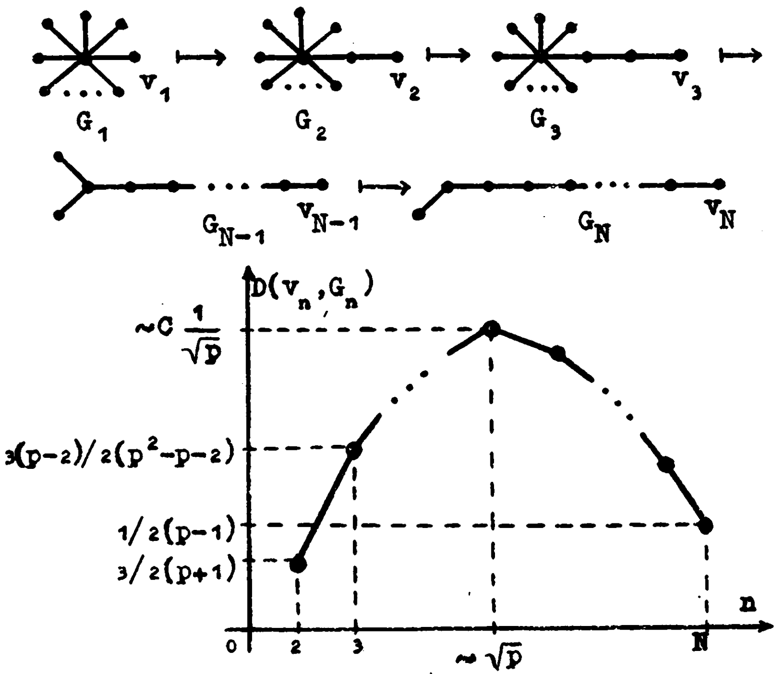



Рис. 7

концевой вершине которой присоединена звезда порядка $p-n$, т.е. графы в парах $G_i$ и $G_{i+1}$ имеют сходную структуру. Для любого $n$ вершина $v_n$ является диаметральной в $G_n$ и имеет максимальную дистанцию среди всех других вершин графа.Характер изменения значений отношения $D(v_n)/D(G_n)$ при $n \to p$ графически представлен на рис. 7. Точные выражения для значений дистанций всех вершин в графах $G_i$ и дистанции $G_i$ как функций от диаметра и порядка графа приводятся в приложении.

Для графов достаточно простой структуры можно в явном виде установить зависимости между дистанционными и эксцентриситетными характеристиками:

1) для полного графа $D(K_p) = (p-1)e(K_p)$;

2) для звезды $K_{1,p-1}$ имеем $D(K_{1,p-1}) = 2p^2 - 2e(K_{1,p-1})$;

3) для колеса $W_p$ выполняется $D(W_p) = 2(p^2-p+1) - 2e(W_p)$;

4) для уникально-эксцентриситетных гладких графов [16]

$$D(G) = \frac{1}{2} pe(G);$$

5) для простого цикла $C_p$ имеем

$$D(C_p) = \begin{cases} \frac{1}{2} pe(G), & p \text{ четно}, \\ \frac{1}{2} \frac{p^2-1}{p} e(G), & p \text{ нечетно}; \end{cases}$$

6) для регулярных графов степени $r$ и диаметра 2

$$D(G) = \frac{1}{2} (2p-r-2)e(G);$$

7) для простой цепи $P$ порядка $p$

$$D(P) = \begin{cases} \frac{4}{3} \frac{p(p+1)}{3p+1} \cdot e(P), & p \text{ четно}, \\ \frac{4}{3} \frac{p^2-1}{3p-2} \cdot e(P), & p \text{ нечетно}. \end{cases}$$

Для цепных инвариантов $D_\tau(v)/D_\tau(G)$ и $e_\tau(v)/e_\tau(G)$ одновершинного фрагмента в классе деревьев их значения совпадают с соответствующими метрическими характеристиками. Вычисление значений цепных характеристик вершин в явном виде для графов сложной структуры существенно более трудоемко из-за комбинаторного роста числа цепей различной длины в графе. Это удается сделать для графов с "обозримой" структурой цепей. Для графов полициклических систем на рис.8, один из которых принадле-

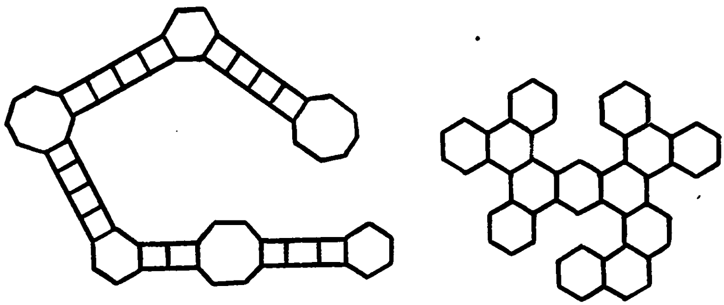


Рис. 8

жит классу полициклических ароматических углеводородов, для любой вершины $v$ выполняется $e_\tau(v)/e_\tau(G) = 1/p$ , так как гамильтонову цепь из каждой вершины можно провести по периметру графа. Графы полициклических систем указанного типа дают пример графов, на которых вырождаются цепные эксцентриситетные инварианты, в то время как метрические эксцентриситетные инварианты успешно разделяют вершины графов.

## §5. Взаимное положение подграфов

Естественным обобщением метрических и цепных инвариантов, описывающих положение подграфа в графе, являются инварианты, предназначенные для определения положения подграфа относительно другого подграфа.

*Дистанцией между подграфами* $F, H \subseteq G$ называется функция

$$D_G(F,H) = \frac{1}{D(G)} \sum_{\substack{v \in V(F) \\ u \in V(H)}} d_G(v,u).$$

*Эксцентриситет между подграфами* $F, H \subseteq G$ задается выражением

$$e_G(F,H) = \frac{1}{e(G)} \sum_{\substack{v \in V(F) \\ u \in V(H)}} \max d_G(v,u).$$

Другие рассматриваемые метрические и цепные инварианты определяются аналогично. При совпадении фрагмента $H$ с графом $G$ эти инварианты совпадают с ранее введенными функциями, описывающими положение подграфа в графе. Такие инварианты будем называть метрическими и цепными инвариантами взаимного положения подграфов в графе. Вместо вершин подграфа $H$ в формулу можно подставить любое другое подмножество вершин $V' \subseteq V(G)$. Тогда считаем, что инвариант $D_G(F, \langle V' \rangle)$ характеризует взаимное положение подграфов $F$ и $\langle V' \rangle$ в графе $G$. Из симметричности расстояний следует выполнение $D(F,H) = D(H,F)$ для произвольных подграфов. Предложенные инварианты учитывают возможность пересечения подграфов по вершинам и ребрам. В табл.3 приводятся примеры фрагментов и значения их взаимного положения.

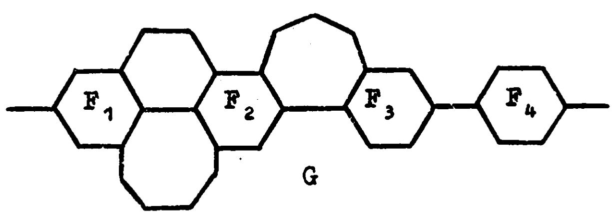
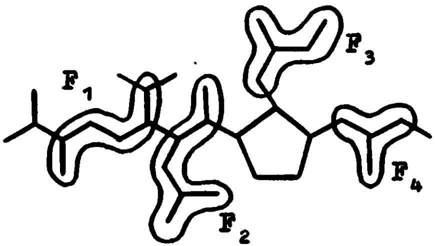
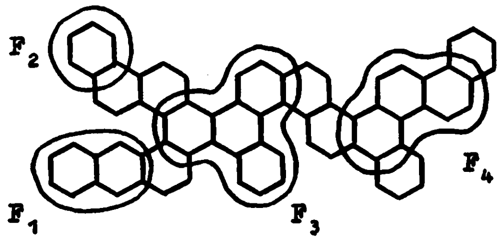



Рис. 9

Т а б л и ц а 3

Метрические инварианты взаимного положения подграфов в молекулярных графах




| Функции | $F_1,F_2$ | $F_1,F_3$ | $F_1,F_4$ | $F_2,F_3$ | $F_1,F_3$ | $F_2,F_3$ | $F_3,F_4$ | $F_1,F_4$ | $F_2,F_4$ |
|---|---|---|---|---|---|---|---|---|---|
| $D(F_i,F_j)$ | 0.0349 | 0.0401 | 0.0346 | 0.0355 | 0.0316 | 0.0215 | 0.0549 | 0.0517 | 0.0310 |
| $D_\tau(F_i,F_j)$ | 0.0178 | 0.0482 | 0.0372 | 0.0508 | 0.4579 | 0.3177 | 0.2748 | 0.1024 | 0.0830 |
| $e(F_i,F_j)$ | 0.0975 | 0.1643 | 0.1643 | 0.1727 | 0.0857 | 0.0563 | 0.1788 | 0.1510 | 0.0906 |
| $e_\tau(F_i,F_j)$ | 0.0827 | 0.1820 | 0.1537 | 0.2033 | 0.1400 | 0.0843 | 0.1928 | 0.1377 | 0.0818 |

Для пар подграфов из набора шестичленных циклов $F_1$, $F_2, F_3, F_4 \subseteq G$ в графе на рис. 9 их взаимные положения совпадают: $D(F_1,F_2) = D(F_3,F_4) = 0.0245$, $D(F_1,F_3) = D(F_2,F_4) = 0.0490$ и $D(F_1,F_4) \neq D(F_2,F_3)$, а значения цепных инвариантов различаются: $D_\tau(F_1,F_2) = 0.0163$, $D_\tau(F_3,F_4) = 0.0166$, $D_\tau(F_1,F_3) = 0.0389$ и $D_\tau(F_2,F_4) = 0.0563$. Поведение эксцентриситетных инвариантов аналогично дистанционным: $e(F_1,F_2) = e(F_3,F_4) = 0.0880$, $e(F_1,F_3) = e(F_2,F_4) = 0.1520$, для цепных выполняются $e_\tau(F_1,F_2) = 0.1568$, $e_\tau(F_3,F_4) = 0.1600$, $e_\tau(F_1,F_3) = 0.1526$, $e_\tau(F_2,F_4) = 0.1768$.

Для метрической характеризации взаимного положения семейства подграфов $F_1,F_2,\ldots,F_n \subseteq G$ относительно друг друга в графе $G$ используем метрическую функцию аналогичного вида

$$D(F_1,F_2,\ldots,F_n) = \frac{1}{2n}\sum_{i,j=1}^{n} D(F_i,F_j),$$

являющуюся также симметрической. Взаимное положение троек подграфов из набора подграфов на рис. 9 принимает значения $D(F_1,F_2,F_3) = D(F_2,F_3,F_4) = 0.0327$ и $D(F_1,F_2,F_4) = D(F_4,F_3,F_1) = 0.0490$, цепные инварианты различают взаимные положения фрагментов $D_\tau(F_1,F_2,F_3) = 0.0258$, $D_\tau(F_2,F_3,F_4) = 0.0317$, $D_\tau(F_1,F_2,F_4) = 0.0572$ и $D\ (F_4,F_3,F_1) = 0.0515$.

## §6. Положение графа $F$ как подграфа в $G$

В предыдущих параграфах при метрической характеризации фрагментов графа считалось, что вершины фрагмента заданы. В случае, когда граф $F$ рассматривается как подграф графа $G$ и соответствие вершин графов не известно, задача метрической характеризации усложняется. Граф $F$ *изоморфно вкладывается* в граф $G$, если существует порожденный подграф $F_1 \subseteq G$ такой, что $F \cong F_1$. Подграф $F_1$ называется *вложением* графа $F$ в граф $G$, которое полностью определяется соответствием вершин

$F_1$ и $F$, называемым подстановкой вложения $F$ в $G$. Тогда задачу можно сформулировать в следующем виде: для заданных графов $F$ и $G$ необходимо характеризовать положение графа $F$ в графе $G$. Вообще говоря, для графа $F$ может существовать несколько вложений, различающихся метрическими свойствами. Вложения $F_1'$ и $F_2$ графа $F$ в $G$ называются *автоморфными*, если существует элемент группы автоморфизмов графа $G$, переводящий подграф $F_1$ в подграф $F_2$. Автоморфные вложения являются полностью эквивалентными по своим структурным и метрическим свойствам. Для неавтоморфных вложений метрические свойства тоже могут совпадать. Например, для неавтоморфных вложений $F_1$ и $F_4$, $F_2$ и $F_3$ шестичленного кольца на рис.9 их дистанции в графе совпадают. Полная информация о вложениях графа $F$ в граф $G$ содержится в двух наборах $N = (F_1, F_2, \ldots, F_n)$ и $M = (m_1, m_2, \ldots, m_n)$, где $F_i$, $i = 1,2,\ldots, n$, - попарно неавтоморфные вложения $F$ в $G$,

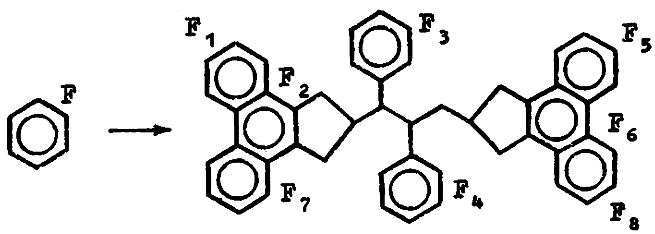



Рис. 10

$m_i$ - кратность вложения $F_i$, т.е. число автоморфных $F_i$ вложений $F$ в граф. Для шестичленного кольца на рис. 10 все варианты его вложения в граф $G$ определяются наборами $N = (F_1, F_2, F_3, F_4, F_5, F_6)$ и $M = (2,1,1,1,2,1)$, а вложения $F_1$ и $F_7$, $F_5$ и $F_8$ являются автоморфными. Определив значения инвариантов для всех вложений, положение $F$ в графе $G$ можно характеризовать средним, максимальным или минимальным значениями метрических инвариантов вложений из набора $N$, интервалом $[f_{min}, f_{max}]$ и другими аналогичными параметрами.

## §7. Положение подграфа в семействе графов

Рассмотрим характеризацию положения графа $F$ в семействе графов $G_i$, $i = 1,2,\dots,n$. Считаем, что для каждого графа семейства задано некоторое вложение в него графа $F$. Для оценки положения графа $F$ необходимо использовать такие метрические инварианты, значения которых для существенно различных положений $F$ в графах $G_i$ не являются близкими. Использование для этих целей рассмотренных ранее метрических характеристик оказывается неприемлемым вследствие сильного влияния возможных различий параметров графов (числа вершин и т.п.) в семействе на значения этих характеристик. Для примера рассмотрим графы $G_1$ и $G_2$ на рис.11. Метрические характеристики

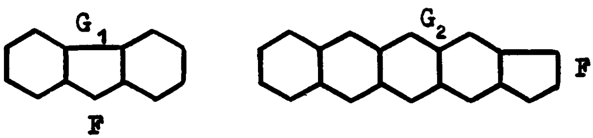



Рис. 11

$D(F,G_1)$ и $D(F,G_2)$ пятичленного кольца $F$ имеют близкие значения, хотя интуитивно ясно, что подграф $F$ занимает различные положения в графах. Действительно, $D(F,G_1) = 0.3151$ и $D(F,G_2) = 0.2765$, а возрастание дистанции подграфа $F$ в $G_2$ компенсируется возрастанием дистанции самого графа $G_2$. Обозначим $m_k(G) = \min_S \{ \sum_{v\in S} D_G(v), S \subseteq V(G), |S| = k\}$ и $M_k(G) = \max_S \{ \sum_{v \in S} D_G(v), S \subseteq V(G), |S| = k\}$. Величины $m_k(G)$ и $M_k(G)$ являются границами для значения дистанции произвольного подграфа порядка $k$ в графе $G$, т.е. для любого $F \subseteq G$, $|V(F)| = k$, выполняется $m_k(G) \le D_G(F) \le M_k(G)$. Множества вершин графа, на которых достигается $m_k(G)$ для дистанций и эксцентриситетов, известны под названием $k$-центра и $k$-центроида графа [45]. Таким образом, положение подграфа $F$ можно характеризовать отклонением его дистанции от границ $m_k(G)$ и $M_k(G)$. Возможны следующие варианты перебора подграфов на множестве вершин $S$ графа для вычисления величин $M_k(G)$ и $m_k(G)$ :

1) перебор всех $k$-элементных подмножеств $S \subset V(G)$, подграфы $\langle S \rangle$ могут быть несвязными;

2) перебор всех связных $k$-вершинных фрагментов $\langle S \rangle$ графа;

3) если фрагмент имеет кратные ребра, то перебираются все скелеты фрагмента $\langle S \rangle$ (кратные ребра не учитываются);

4) если фрагмент имеет метки вершин, то перебираются все непомеченные фрагменты $\langle S \rangle$ (метки вершин не учитываются);

5) перебор всех точных вложений помеченного по ребрам и вершинам фрагмента $\langle S \rangle$ в граф;

6) перебор частичных подграфов на подмножествах $S \subset V(G)$ с возможным ограничением на количество ребер частичного подграфа и т.п.

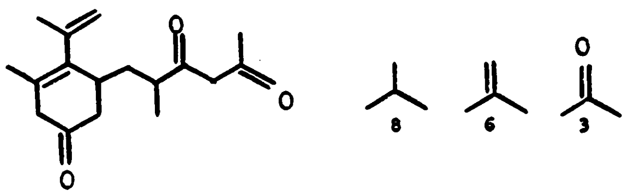



Рис. 12

Для фрагментов на рис. 12 приводятся их количества в графе без учета меток вершин и кратностей ребер, с учетом кратностей ребер и при точном вложении подграфов.

Положение подграфа $F$ в $G$ будем называть центральным, если $D_G(F) = m_k(G)$ и периферийным при $D_G(F) = M_k(G)$. Для оценки положения подграфа $F$ порядка $k$ в графе $G$ будем использовать величину

$$R_G(F) = \frac{D_G(F)-m_k(G)}{M_k(G)-m_k(G)} .$$

Значение $R_G(F)$ есть нормированное отклонение дистанции подграфа $F$ от значений $m_k(G)$ и $M_k(G)$ и $0 \le R_G(F) \le 1$. В выражении для $R_G(F)$ разность $M_k(G) - m_k(G)$ принимает нулевое значение, когда графы $F$ и $G$ имеют одинаковые порядки или при равенстве $D(u) = D(v)$ для любых вершин $u,v \in V(G)$. В этом случае считаем, что положение подграфа $F$ является центральным в $G$. Окончательно $R_G(F)$ примет следующий вид:

$$R_G(F)= \begin{cases} \dfrac{D_G(F)-m_k(G)}{M_k(G)-m_k(G)}, & \text{если } k \neq p \text{ и существуют вершины } u,v \in V(G) \text{ такие, что } D(u) \neq D(v), \\ 0 - & \text{в противном случае,} \end{cases}$$

где $p$ - порядок графа $G$, $k$ - порядок графа $F$. Значения $R_{G_i}(F)$, $i = 1,2,\ldots,n$, не зависят от различий параметров в графах семейства и их можно использовать для метрической характеризации положения графа $F$ в семействе произвольных графов $G_1,G_2,\ldots,G_n$. Так, для рассматриваемого нами пятичленного кольца в графах на рис.11 выполняются $R_{G_1}(F) = 0.00$ и $R_{G_2}(F) = 0.76$.

## §8. Метрические инварианты частичных подграфов

Задача характеризации частичных подграфов имеет особенность, так как частичные подграфы на одном и том же множестве вершин различаются количеством принадлежащих им ребер. Ниже предлагается способ метрической характеризации частичных подграфов, использующий понятие метрического инварианта ребра. Пусть $y=(v,u)$ есть ребро графа $G$. Определим дистанцию ребра $y$ в графе как величину $D'(y) = \frac{1}{2}(D(v)+D(u))$.

На основе дистанции ребра введем реберную дистанцию графа, задаваемую выражением

$$D'(G) = \sum_{y\in E(G)} D'(y) = \frac{1}{2}\sum_{(v,u)\in E(G)} (D(v)+D(u)) =$$

$$= \frac{1}{2}\sum_{v\in V(G)} \deg(v)D(v),$$

где $\deg(v)$ есть степень вершины $v$ в графе. Из последнего значения видно, как возрастает реберная дистанция графа по сравнению с вершинной дистанцией графа. Например, для $r$-регулярного графа выполняется $rD(G) = 2D'(G)$, а для графов с совпадающими значениями дистанций для всех вершин имеем $D'(G) = qD(v)$. Используя утверждение 2, можно получить границы значений отношения реберной и вершинной дистанций графа $\frac{q}{2p} \le \frac{D'(G)}{D(G)} \le \sqrt{2}\frac{q}{\sqrt{p}}$. Реберный эксцентриситет графа определяется аналогично

$$e'(G) = \sum_{(v,u)\in E(G)} \frac{1}{2}(e(v)+e(u)) =$$

$$= \frac{1}{2}\sum_{v\in V(G)} \deg(v)e(v).$$

Очевидны неравенства $r(G)q \le e'(G) \le d(G)q$, а для вершинного эксцентриситета графа $r(G)p \le e(G) \le d(G)p$. Для отношений эксцентриситетов графа по утверждению 2 имеем $\frac{q}{2p-1} \le \frac{e'(G)}{e(G)} \le \frac{2q}{p+2}$.

Для частичного подграфа $F \subseteq G$ его реберная дистанция

$$D'_G(F) = \sum_{y \in E(F)} D'(y) = \frac{1}{2} \sum_{(v,u) \in E(F)} (D_G(v) + D_G(u)) =$$

$$= \frac{1}{2} \sum_{v \in V(F)} \deg_F(v) D_G(v),$$

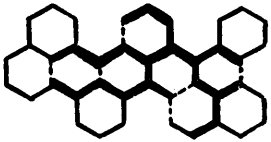
где $\deg_F(v)$ есть число ребер подграфа, смежных с вершиной $v$. Нормированная реберная дистанция частичного подграфа $F$ определяется как $D'(F,G) = D'_G(F)/D'(G)$. Для эксцентриситетов подграфа $e'_G(F) = \frac{1}{2} \sum_{v \in V(F)} \deg_F(v) e_G(v)$ и $e'(F,G) = e'_G(F)/e'(G)$.

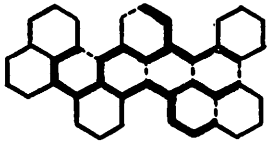
Для изолированной вершины $v$ подграфа можно условно считать, что она содержит петлю $y = (v,v)$, тогда $D'(y) = \frac{1}{2}(D(v) + D(v)) = D(v)$, т.е. изолированная вершина характеризуется своей обычной дистанцией. Если по вкладу в значение инварианта изолированная вершина должна существенно отличаться от ребра, то можно положить $D(y) = \frac{1}{2} D(v)$ и т.д.

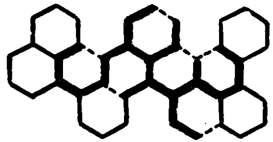
Таким же образом на основе реберных инвариантов строятся цепные характеристики графов и частичных подграфов. В табл. 4 приводятся значения метрических и цепных инвариантов для некоторых частичных подграфов на одном и том же подмножестве вершин полициклического молекулярного графа. Для порожденного подграфа также приведены значения его реберных инвариантов.

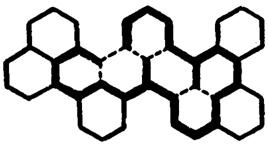
Т а б л и ц а 4

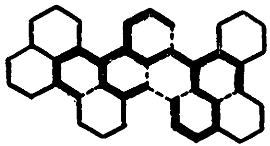
Метрические инварианты частичных подграфов молекулярного графа

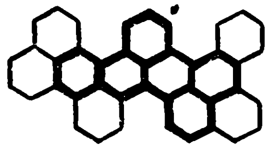
$F_1$


$F_2$


$F_3$


$F_4$


$F_5$


$F_6$


| Функции | $F_1$ | $F_2$ | $F_3$ | $F_4$ | $F_5$ | $F_6$ |
|---|---|---|---|---|---|---|
| $D'(F,G)$ | 0.3970 | 0.3964 | 0.3977 | 0.4082 | 0.4040 | 0.5028 |
| $D'_\tau(F,G)$ | 0.4058 | 0.4093 | 0.4091 | 0.4140 | 0.4158 | 0.5152 |
| $e'(F,G)$ | 0.4010 | 0.4029 | 0.4039 | 0.4126 | 0.4106 | 0.5101 |
| $e'_\tau(F,G)$ | 0.4465 | 0.4475 | 0.4475 | 0.4475 | 0.4485 | 0.5640 |

## §9. Другие схемы вычисления инвариантов фрагментов

В этом параграфе рассматриваются другие возможные схемы вычисления метрических инвариантов фрагментов графа.

Схема Мекеняна-Бончева-Балабана. В [17] предложена схема построения топологических индексов для характеризации фрагментов молекулярных графов. Изложим основные идеи этого подхода. Для фрагмента $F \subsetneq G$ определяется "внутренний" топологический индекс (Internal Fragment Topological Index, IFTI), описывающий топологию фрагмента и представляющий собой интегральный топологический индекс для $F$ как отдельного графа. Считается, что фрагмент $F$ является всегда порожденным подграфом. Для описания связи фрагмента с оставшейся частью молекулярного графа используется "внешний" топологический индекс фрагмента (External Fragment Topological Index, EFTI), задаваемый выражением $\mathrm{EFTI}(F) = \mathrm{TI}(G) - [\mathrm{IFTI}(F) - \sum_k \mathrm{IFTI}\,(G-F)_k]$,

где $\mathrm{TI}(G)$ есть топологический индекс графа $G$. Суммирование ведется по всем связным компонентам подграфа $G-F$, получаемого удалением из $G$ всех вершин фрагмента $F$ и инцидентных им ребер (рис. 13). Предполагается, что $\mathrm{TI}(G)$ является возрастающей функцией при увеличении числа вершин графа. Для индексов фрагментов формулируются естественные требования

$$0 \le \mathrm{IFTI}(F) \le \mathrm{TI}(G), \qquad (1)$$
$$0 \le \mathrm{IFTI}(G-F) \le \mathrm{TI}(G).$$

Для EFTI индекса выполняется $\mathrm{EFTI} < \mathrm{TI}(G)$. Так как всегда существует связь между вершинами подграфов $F$ и $G-F$ вследствие связности $G$, то $\mathrm{EFTI} > 0$, за исключением случаев, при которых выполняется неравенство

$$\mathrm{IFTI}(F) - \sum_k \mathrm{IFTI}(G-F)_k < \mathrm{TI}(G). \qquad (2)$$

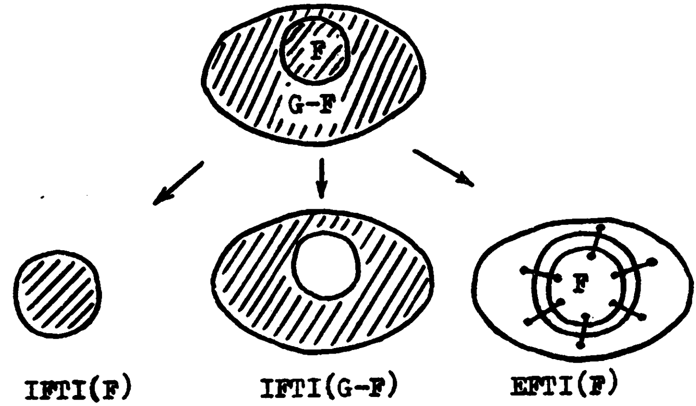



Рис. 13

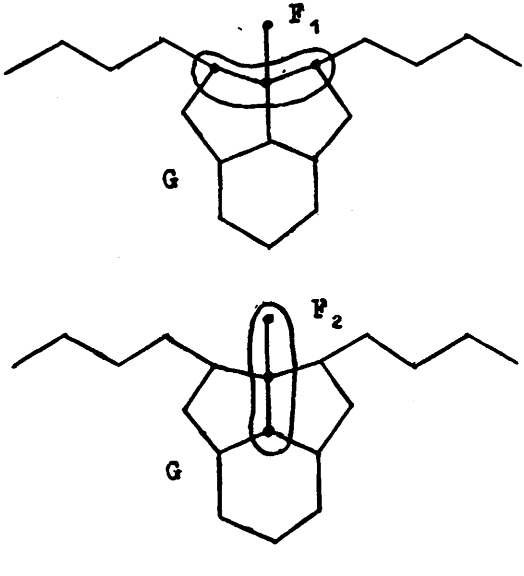



Рис. 14

Нормированные индексы NIFTI и NEFTI образуются отношением значений IFTI и EFTI к индексу всего графа TI(G) .

В рамках этой схемы исследовалось и обсуждалось поведение известных топологических индексов - индекса вершинной связности VA , загребского индекса $M_1$ , индекса молекулярной связности (индекса Рандича) $^1\chi$ , индекса обобщенной связности $^h\chi$ , индекса валентной связности $^h\chi^v$ ,индекса реберной смежности (индекс Гордона-Скантлбери), индекса Винера W , информационного аналога индекса Винера $I_D^M$ , индекса Хосои Z , индексов иерархической упорядоченной расширенной связности НОС (индексы Бончева-Балабана-Мекеняна), индекса Балабана J и др. [6,7,9,11,12]. Построенные на основе индекса Винера $W(G)= \frac{1}{2} \sum\limits_{v,u \in V(G)} d(v,u)$ индексы для F имеют вид:

$$IFTI.W(F) = \frac{1}{2} \sum_{v,u \in V(F)} d_F(v,u),$$

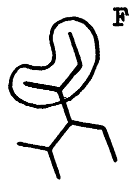
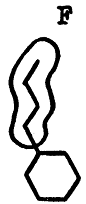
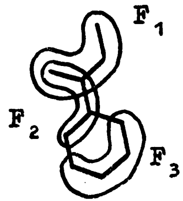
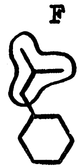
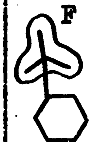
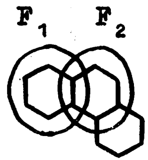
$$EFTI.W(F) = W(G)-[IFTI.W(F)- \sum_k (IFTI.W(G-F)_k].$$

В табл. 5 для сравнения приводятся численные значения индекса Винера фрагментов в схеме Мекеняна-Бончева-Балабана и значения рассматриваемых метрических инвариантов в предлагаемой нами схеме.

Отметим одну особенность характеризации фрагментов с использованием индекса Винера по схеме Мекеняна-Бончева-Балабана, не отраженную в [17]. Пусть фрагмент F расположен в графе G несколькими способами, тогда оказывается, что для одного положения фрагмента F неравенства (1) выполняются, а для другого положения F эти неравенства изменяются на противопо-

Таблица 5

Метрические инварианты подграфов для разных схем вычислений

|  |  |  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|
| Функции | F | F | $F_1$ | $F_2$ | $F_3$ | F | F | $F_1$ | $F_2$ |
| D(F,G) | 0.4098 | 0.4361 | 0.4298 | 0.3471 | 0.4132 | 0.4286 | 0.4211 | 0.4317 | 0.3653 |
| NIFTI.W | 0.0820 | 0.0752 | 0.0826 | 0.0826 | 0.0826 | 0.0714 | 0.0789 | 0.0996 | 0.0996 |
| NEFTI.W | 0.6557 | 0.7218 | 0.6942 | 0.6612 | 0.8264 | 0.7143 | 0.6842 | 0.6642 | 0.8266 |

ложные. Для иллюстрации рассмотрим граф $G$ на рис. 14 и два его фрагмента $F_1 \cong F_2$, являющихся простыми цепями на трех вершинах. Для индексов фрагментов имеем $IFTI.W(F_1) = IFTI.W(F_2) = 4$, $IFTI.W(G-F_1) = 81$ и $IFTI.W(G-F_2) = 816$, для всего графа индекс Винера $W(G) = 748$. Знак "внешних" индексов $EFTI.W(F_1) = 663$ и $EFTI.W(F_2) = -72$ меняется на противоположный, т.е. не совсем ясно, как в данной вычислительной схеме применять индексы типа индекса Винера для характеризации фрагментов. Наблюдаемое колебание значений индекса Винера для фрагментов следует из того факта, что для расстояния между вершинами $v,u \in V(F)$ произвольного подграфа $F \subseteq G$ выполняется неравенство $d_G(v,u) \leq d_F(v,u)$, которое влечет выполнение $\sum_{v,u \in V(F)} d_G(v,u) \leq \sum_{v,u \in V(F)} d_F(v,u)$. Действительно, фрагмент $F$ можно получить из графа $G$ удалением лишних вершин и ребер, при этом расстояния между оставшимися вершинами могут только увеличиваться. Все расстояния между вершинами фрагмента $F \subseteq G$, рассматриваемого как граф, и теми же вершинами в графе $G$ совпадают, если $G$ является деревом, если $F$ расположен в "древесной" части $G$ или целиком включает циклы, к которым принадлежит хотя бы одна вершина $F$, если связные части цикла, входящего в $F$, имеют длину менее половины длины соответствующего цикла и т.п. Интересный класс графов образуют наследственные по расстоянию графы, в которых для любого порожденного подграфа расстояние между его произвольной парой вершин совпадает с расстоянием между этими вершинами в графе. Отсюда следует, что циклы длины более 4 в таких графах должны иметь диагонали и т.п. Полная характеризация наследственных по расстоянию графов дана в [46].

Предложенную нами в предыдущих параграфах схему определения инвариантов фрагментов можно представить в виде рассматриваемой в этом параграфе схемы. Действительно, для дистанции графа можно записать

$$D(G) = D_G(F)+D_G(G-F) = D(F) +$$

$$+ \sum_{\substack{v\in V(F)\\ u\in V(G-F)}} d_G(v,u)+D(G-F)+ \sum_{\substack{v\in V(G-F)\\ u\in V(F)}} d_G(v,u) =$$

$$= D(F)+D(G-F)+ 2\cdot\sum_{\substack{v\in V(G-F)\\ u\in V(F)}} d_G(v,u).$$

Представим это в виде разности

$$D(G)-[D_G(F)+D_G(G-F)] = 2\cdot\sum_{\substack{v\in V(F)\\ u\in V(G-F)}} d_G(v,u).$$

Левая часть последнего равенства по форме идентична выражению для "внешнего" индекса $EFTI.W(F)$ . Правая часть равенства есть не что иное, как сумма расстояний между вершинами фрагмента и вершинами оставшейся части графа, т.е. является характеристикой связи фрагмента $F$ с $G-F$ . Таким образом, отличие состоит в различном понимании оставшейся части графа. В нашей схеме под остатком графа понимается просто множество вершин $V(G) - V(F)$ в том же графе $G$ , а в рассматриваемой схеме под остатком графа понимается граф $G-F$ , топология которого может сильно отличаться от топологии $G$ .

<u>Схемы с неполным учетом расстояний</u>. Рассмотрим другие возможные схемы вычисления метрических и цепных инвариантов фрагментов графов.

В первой схеме при построении метрических инвариантов для фрагмента $F$ не учитываются расстояния между его вершинами, т.е. $D_G^*(F) = \{\Sigma\, d_G(v,u) \mid v\in V(F),\ u\in V(G)\setminus V(F)\}$. Так как значащими являются только связи между вершинами фрагмента $F$ и не принадлежащими ему вершинами графа $G$, то такой инвариант слабо характеризует топологию самого фрагмента. Покажем, что при увеличении порядка фрагмента значения $D_G^*(F)$ могут

изменяться немонотонно. Рассмотрим простой цикл $C_p$ и его связные фрагменты $F_1 \cong P_m$ и $F_2 \cong P_{p-2m}$ такие, что $F_1 \subset F_2$. Тогда нетрудно заметить, что $D^*_{C_p}(F_1) =$
$= \{\Sigma\, d_{C_p}(v,u) \mid v \in V(F_1),\quad u \in V(C_p) \setminus V(F_1)\} =$
$= \{\Sigma\, d_{C_p}(v,u) \mid v \in V(C_p) \setminus V(F_2),\ u \in V(F_2)\} = D^*_{C_p}(F_2)$.
Два подграфа в $C_p$, один из которых содержит другой, имеют одинаковое значение метрического инварианта $D^*_{C_p}(F)$, т.е. различная топология фрагментов в этом случае никак не сказывается на значении инварианта. В дополнение к $D^*_{C_p}(F)$ топологию фрагмента можно учитывать дистанцией $D(F)$, рассматривая $F$ как граф, и характеризовать положение фрагмента двумя числами. Отметим, что в общем случае для метрических инвариантов выполняется $D(F)+D^*_G(F) \neq D_G(F)$ из-за возможного несовпадения расстояний $d_F(v,u)$ и $d_G(v,u)$ для $v,u \in V(F)$.

Во второй схеме вычислений не учитываются не только расстояния между вершинами фрагмента, но и структура самого фрагмента. Все вершины связного подграфа стягиваются в одну вершину с удалением петель и кратных ребер, и далее характеризуется положение такого одновершинного фрагмента. В случае несвязного подграфа этот процесс производится для каждой связной компоненты, и результатом является значение метрического инварианта на полученном вполне несвязном подграфе. При таком подходе может сильно измениться структура исходного графа, и метрическими инвариантами будет характеризоваться модифицированный фрагмент в редуцированном графе, топологию же исходного фрагмента можно дополнительно описывать значением его дистанции $D(F)$.

Схемы с неполным учетом расстояний представляются менее перспективными для использования с целью метрической характеризации фрагментов.

## §10. Вычисление инвариантов фрагментов

Опишем способы вычисления рассмотренных выше метрических и цепных инвариантов фрагментов графа. Метрические и цепные инварианты единообразно вычисляются по матрицам слоев графа, известным также как дистанционная и цепная степенные последовательности графа [47].

В матрице слоев $\lambda(G) = \|\lambda_{ij}\|$, $i = 1,2,\ldots, p$, $j = 1,2,\ldots, p-1$, элемент $\lambda_{ij}$ есть число вершин, находящихся на расстоянии $j$ от вершины $v_i$ в графе $G$. Метрические инварианты графов вычисляются также и по матрице расстояний графа $MD(G)$ [5,6,8,9,11]. С матрицей расстояний графа матрица $\lambda(G)$ связана следующим образом: элемент $\lambda_{ij}$ равен числу компонент равных $j$ в строке для вершины $v_i$ в матрице расстояний. Матрицы смежностей, расстояний и матрица слоев графа и его фрагмента приведены на рис. 15.

Число ненулевых элементов в строке $\lambda(G)$ равно эксцентриситету соответствующей вершины. Метрические инварианты для подграфа $F \subseteq G$ вычисляются по матрице слоев как

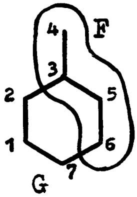
$$D(v_i) = \sum_{j=1}^{p-1} j\lambda_{ij}, \quad D(G) = \sum_{i=1}^{p} \sum_{j=1}^{p-1} j\lambda_{ij},$$

$$D(F) = \sum_{v_i \in V(F)} \left( \sum_{j=1}^{p-1} j\lambda_{ij} \right)$$

и

$$e(G) = \sum_{i=1}^{p} e(v_i), \quad e(F) = \sum_{v_i \in V(F)} e(v_i).$$

Вырождаемость матрицы слоев для пары неизоморфных графов влечет и вырождаемость всех построенных по ней метрических инвариантов. Вершины кубического графа с $p = 18$ на рис.16 имеют совпадающие значения дистанции, так как все строки матрицы сло-




| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 2 | 2 | 2 | 0 |
| 2 | 2 | 3 | 1 | 0 |
| 3 | 3 | 2 | 1 | 0 |
| 4 | 1 | 2 | 2 | 1 |
| 5 | 2 | 3 | 1 | 0 |
| 6 | 2 | 2 | 2 | 0 |
| 7 | 2 | 2 | 1 | 1 |

$\lambda(G)$

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 2 | 2 | 3 | 2 | 3 | 0 |
| 2 | 2 | 3 | 2 | 2 | 2 | 1 |
| 3 | 3 | 2 | 2 | 2 | 2 | 0 |
| 4 | 1 | 2 | 2 | 2 | 2 | 2 |
| 5 | 2 | 3 | 2 | 2 | 2 | 1 |
| 6 | 2 | 2 | 3 | 2 | 3 | 0 |
| 7 | 2 | 2 | 2 | 4 | 2 | 0 |

$\tau(G)$

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | · | 1 | · | · | · | · | 1 |
| 2 | 1 | · | 1 | · | · | · | · |
| 3 | · | 1 | · | 1 | 1 | · | · |
| 4 | · | · | 1 | · | · | · | · |
| 5 | · | · | 1 | · | · | 1 | · |
| 6 | · | · | · | · | 1 | · | 1 |
| 7 | 1 | · | · | · | · | 1 | · |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 2 | 3 | 3 | 2 | 1 |
| 2 | 1 | 0 | 1 | 2 | 2 | 3 | 2 |
| 3 | 2 | 1 | 0 | 1 | 1 | 2 | 3 |
| 4 | 3 | 2 | 1 | 2 | 0 | 1 | 4 |
| 5 | 3 | 2 | 1 | 2 | 0 | 1 | 2 |
| 6 | 2 | 3 | 2 | 3 | 1 | 0 | 1 |
| 7 | 1 | 2 | 3 | 4 | 2 | 1 | 0 |

Рис. 15

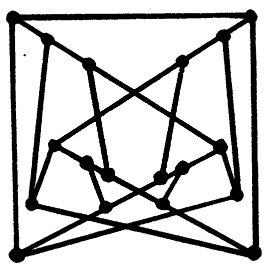


Рис. 16

ев графа совпадают и равны (3,6,7,1) (вершины содержатся в 9 орбитах мощности 2 группы автоморфизмов графа). Проблема вырождаемости матрицы $\lambda(G)$ (однозначности представления графов) изучалась в [21,23,47-51]. Отметим, что наименьший известный порядок деревьев с совпадающей матрицей слоев равен 18, а для произвольных графов равен 5 (см. рис. 17а). Матрица $\lambda(G)$ вычисляется по матрице смежности графа с трудоемкостью $O(pq)$ операций простым алгоритмом подразбиения множества всех вершин графа по расстоянию до выделенной вершины.

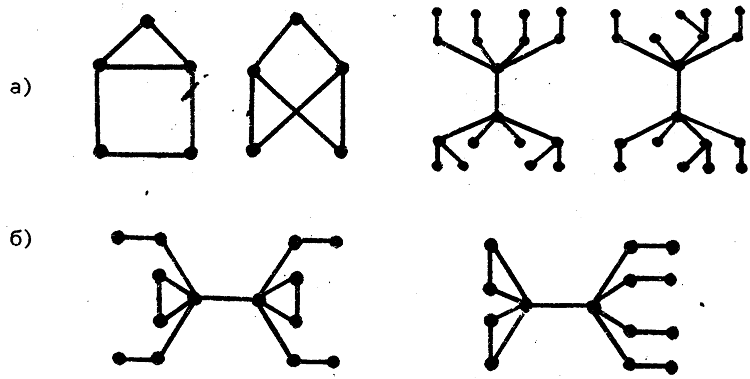



Рис. 17

Инварианты, построенные на основе цепного расстояния и протяженности, вычисляются по цепной матрице слоев $\tau(G) = \|\tau_{ij}\|$, $i = 1,2,\ldots,p$, $j = 1,2,\ldots,p-1$, элемент $\tau_{ij}$ которой равен числу всех простых цепей длины $j$, выходящих из вершины $v_i$. Очевидно, что в классе деревьев выполняется равенство $\tau(G) = \lambda(G)$. По определению $\tau(G)$, число ненулевых элементов в строке равно цепному эксцентриситету (числу протяженности) соответствующей вершины. Через компоненты матрицы $\tau(G)$ цепные характеристики выражаются аналогично рассмотренным выше метрическим инвариантам, например, $D_\tau(G) = \sum_{i=1}^{p} \sum_{j=1}^{p-1} j\tau_{ij}$ и т.п. Цепная матрица слоев более полно отражает структуру графов и имеет меньшую степень вырождаемости по сравнению с матрицей слоев $\lambda(G)$. Однозначность представления графов цепной матрицей слоев изучалась в [23,47, 52]. Наименьший известный порядок неизоморфных графов с совпадающей цепной матрицей слоев равен 14 (рис.17б), в классе кубических графов наименьший порядок таких графов равен 62. В [52] для любых $r \geq 3$ и $n \geq 2$, предложен способ конструктивного построения $n$ попарно неизоморфных $r$-регулярных графов с совпадающими цепными матрицами слоев. Для графов произвольной структуры вычисление цепных инвариантов (цепной матрицы слоев) требует значительных временных затрат, так как задача конструктивного перечисления всех цепей подграфа различной длины имеет переборный характер и алгоритмически решается путем просмотра графа в глубину с возвратом. Однако на практике для некоторых классов графов вычисление цепных характеристик имеет невысокую трудоемкость. К таким классам относится большинство классов графов молекулярных структур. Это объясняется особенностями их структуры - малая степень вершин (обычно $\leq 4$), достаточно большой диаметр по сравнению с числом вершин, планарность, наличие мостов и точек сочленения ("древовидные" части графов) и т.п. Все эти признаки ограничивают рост числа простых цепей в графе. Конечно, и среди графов молекулярных структур встречаются графы, для которых перечисление их цепей весьма трудоемко. Такими являются классы полициклических конденсированных структур с большим числом колец. На рис.18 приведен пример графа из класса гексагональных бензоидных ароматических систем.

В заключение рассмотрим трудоемкость вычисления величин $m_k(G)$ и $M_k(G)$ (см. §7), являющихся минимальным и макси-

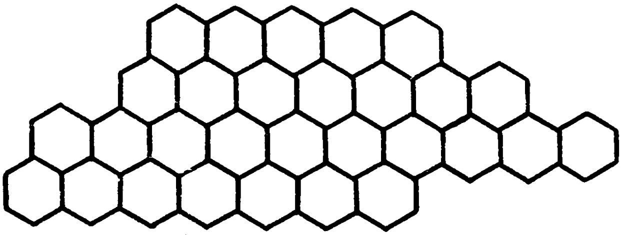


Рис. 18

мальным значениями метрического инварианта на $k$-вершинных подграфах графа $G$. Если эти значения допускается искать на несвязных графах, то трудоемкость вычислений совпадает со сложностью сортировки $p$ чисел и пропорциональна $p \cdot \log_2 p$. Действительно, пусть значения дистанции $D(v)$ для всех вершин $v \in V(G)$ отсортированы в неубывающем порядке. Тогда, суммируя первые $k$ значений, получим величину $m_k(G)$, а сумма последних $k$ элементов даст $M_k(G)$. В других вариантах выбора подграфов трудоемкость существенно возрастает, так как необходимо выделение всех связных подграфов графа, нахождение всех вложений помеченного по вершинам и ребрам фрагмента или его скелета в граф.

## §11. Использование инвариантов фрагментов

Рассмотрим одну из возможностей использования метрических и цепных инвариантов фрагментов при решении задачи установления взаимозависимостей между структурой и свойствами химических соединений в рамках подхода [53,54], основанного на рассмотрении общих частей соединений. Пусть $S_1 = \{G_1, G_2, \ldots, G_n\}$ есть класс молекулярных графов соединений, обладающих свойством $P_1$. В качестве потенциальных признаков структурного подобия рассматриваются все попарные пересечения $G_i \cap G_j$,

$1 \le i < j \le n$ (максимальные по числу вершин связные порожденные общие подграфы) графов класса $S_1$. Из полученного набора потенциальных признаков-фрагментов исключаются изоморфные, и формируется таблица объект-признак $T = \|t_{ij}\|$, $i = 1,2,\ldots,n$, $j = 1,2,\ldots,k$, где $k$ есть число отобранных фрагментов. Элемент $t_{ij}$ определяется как $t_{ij} = 1$, если $j$-й признак содержится в $i$-й структуре и $t_{ij} = 0$ иначе, т.е. $T$ есть характеристическая таблица вхождения фрагментов в молекулярные графы класса $S_1$. Далее таблица $T$ анализируется методами распознавания образов [55], с помощью которых определяется набор наиболее информативных признаков-фрагментов, потенциально ответственных за проявления свойства $P_1$. Для прогнозирования строятся логические решающие правила, дающие информацию о том, какие фрагменты должны содержаться или не содержаться в исходном классе молекулярных графов. Для повышения надежности распознавания, наряду с классом $S_1$, рассматривается класс $S_2$ сходных по структуре молекулярных графов. Соединения класса $S_2$ обладают свойством $P_2$, отличным от $P_1$ по степени проявления, по механизму действия или имеющим другую природу. Между графами класса $S_2$ также находятся все попарные пересечения, определяются вложения фрагментов графов класса $S_1$ в графы класса $S_2$ и наоборот, и далее формируется общая таблица $T$ объект-признак для $S_1$ и $S_2$. Таблица $T$ является качественной для $S_1$ и $S_2$, так как содержит информацию только о наличии или отсутствии фрагментов в графах. Подстановка значений метрических инвариантов фрагментов в таблицу $T$ является естественной попыткой использовать характеризацию положения фрагментов в графах для улучшения результатов. Если фрагмент содержится в графе, то вместо элемента $t_{ij} = 1$ помещается значение метрического инварианта $R_{G_i}(F_j)^{ij}$. Таблица $T$ из качественной преобразуется в количественную. Переход к количественной таблице объект-признак может способствовать повышению информативности признаков-фрагментов. Действительно, предположим, что фрагмент $F_j$ входил во все молекулярные графы классов $S_1$ и $S_2$ (в таблице $T$ $j$-й столбец состоит из единиц). Тогда при анализе качественной таблицы объект-признак этот признак будет отброшен как неинформативный. В количественной таблице метрический инвариант положения фрагментов может иметь значения, которые отделяют молекулярные графы классов $S_1$ и $S_2$ друг от друга.

Для примера рассмотрим молекулярные графы амидинов в двух классах $S_1 = \{1,2,3\}$ и $S_2 = \{4,5\}$ на рис.19. Соединения первого класса являются активными, а соединения второго не проявляют активность. Приведенный на рис. 19 фрагмент $F$ имеется во всех структурах обоих классов, следовательно, учет его вхождения не дает информативного признака. Вычисленные значения инварианта $R_{G_i}(F)$ для фрагмента приводятся в табл.6, где слева указаны функции, подставляемые в формулу инварианта.

Таблица 6

Положение фрагмента в семействе графов

| Функции | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $D_G(F)$ | 0.311 | 0.216 | 0.328 | 0.000 | 0.031 |
| $e_G(F)$ | 0.539 | 0.400 | 0.600 | 0.000 | 0.000 |
| $D_\tau(F)$ | 0.328 | 0.212 | 0.339 | 0.038 | 0.051 |
| $e_\tau(F)$ | 0.500 | 0.400 | 0.455 | 0.143 | 0.167 |

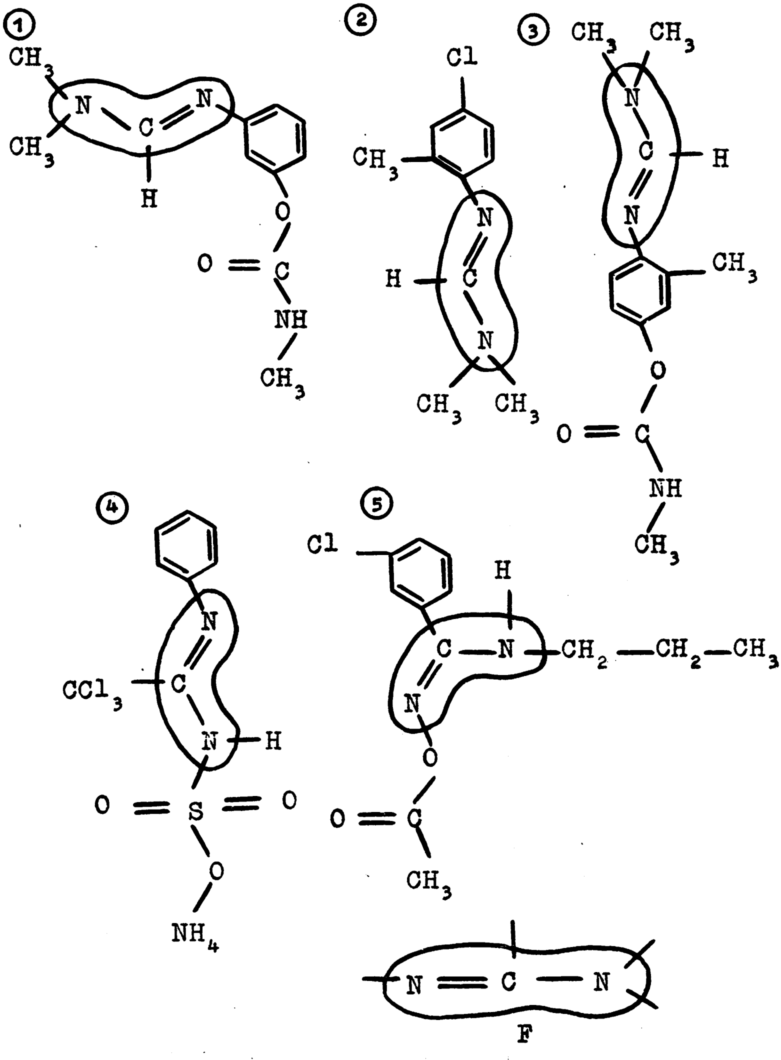



Рис. 19. Положение фрагмента в семействе графов

Из результатов видно, что классы $S_1$ и $S_2$ отделяются по центральному и периферийному положению фрагмента. Следовательно, можно выдвинуть гипотезу о существенности положения данного фрагмента для проявления или степени проявления свойства.

Программные средства вычисления метрических и цепных инвариантов графов и их фрагментов, включая программы нахождения общих частей графов и определения вложения частичных и порожденных, связных и несвязных фрагментов в графы, входят в состав пакета прикладных программ обработки структур молекулярных графов для персонального компьютера [56].

## З а к л ю ч е н и е

В работе предложены подходы к характеризации подграфов графа метрическими и цепными инвариантами,определяемыми на основе различных видов расстояний в графе: естественного расстояния,протяженности и цепного расстояния между вершинами графа. Для инвариантов подграфов сформулированы требования, которые обеспечивают возможность их применения для описания положения подграфа в графе, взаимного положения подграфов в графе и положения подграфа в семействе графов. В качестве инвариантов графов рассматривались дистанционные (дистанция вершины и дистанция графа) и эксцентриситетные (эксцентриситет вершины и эксцентриситет графа) характеристики. Аналогами этих инвариантов являются топологические индексы, используемые для установления взаимосвязей между структурой молекулярных графов и свойствами соответствующих химических соединений. Вычисления инвариантов фрагментов для различных расстояний проводятся по единой методике, основанной на использовании матриц слоев графов. Применение метрических инвариантов распространено на частичные подграфы графов. В работе обсуждались другие схемы построения метрических инвариантов фрагментов графа. В завершение показан возможный путь использования инвариантов фрагмен -

тов в системах прогнозирования свойств, использующих общие части молекулярных графов.

## Литература


1. РОБЕРТС Ф.С. Дискретные математические модели с приложениями к социальным, биологическим и экологическим задачам. - М.: Мир, 1986. - 496 с.

2. ХЕВИК Т., ГЛЕДИЧ Н.П. Структурные параметры графов.Теоретическое исследование //Математика в социологии. - М.: Мир, 1977. - С. 151-169.

3. BALABAN A.T. Applications of graph theory in chemistry //J.Chem. Inf. Comput. Sci., 1985. - Vol. 25. -P. 334-343.

4. Concepts and Applications of Molecular Similatity /Eds. Johnson M.A. and Maggiora G.M.- New York: Wiley, 1990.-393 p.

5. TRINAJSTIĆ N. Chemical graph theory. - Florida: CRC Press, Boca Raton. - 1983. - T. 2.

6. GUTMAN I., POLANSKY O. Mathematical concepts in organic chemistry. - Berlin: Springes-Verlag, 1986.

7. BONCHEV D. Information theoretic indices for characterization of chemical structures. - Chichester: Reseach Studies Press, 1983.

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

9. СТАНКЕВИЧ И.В. Графы в структурной химии //Применение теории графов в химии. - Новосибирск, 1988. - С. 7-69.

10. SABLJIĆ A., TRINAJSTIĆ N. Quantitative structure-activity relationship: the role of topological indices //Acta Pharm. Jugosl. - 1981. - Vol. 34, N 4. -P. 189-414.

11. РУВРЭ Д. Следует ли заниматься разработкой топологических индексов? //Химические приложения топологии и теории графов. - М.: Мир, 1987. - С. 183-205.

12. ROUVRAY D.H. The modeling of chemical phenomena using topological indices //J.Comput. Chem. - 1987.-- Vol. 8, N 4.- P. 470-480.

13. ROUVRAY D.H. The limits of applicability of topological indices //J.Mol. Struct. (Theochem).- 1989. - Vol. 185. - P. 187-201.

14. РУВРЭ Д. Химию прогнозирует топология //В мире науки. - 1986. -№ 11. -С. 14-22.

15. ДРБОГЛАВ В.В. Инварианты графов и их использование для обработки структурной информации: Автореф. дис...канд. техн. наук: 05.13.16. - Новосибирск: 1987. - 16 с.

16. SKOROBOGATOV V.A., DOBRYNIN A.A. Metric analysis of graphs //Comm. Math. Chem. (MATCH). - 1988. - N. 23. - P. 105-151.

17. MEKENYAN O., BONCHEV D. Topological indices for molecular fragments and new graph invariants //J. Math. Chem. - 1988. - N.3. -P. 347-375.

18. ХАРАРИ Ф. Теория графов. - М.: Мир, 1973. - 300 с.

19. ОРЕ О. Теория графов. - М.: Мир, 1982. - 336 с.

20. ДОБРЫНИН А.А., СКОРОБОГАТОВ В.А. Свойства цепей графов и изотопичность //Алгоритмический анализ структурной информации. - Новосибирск, 1985. - Вып. 112: Вычислительные системы. - С. 33-45.

21. СКОРОБОГАТОВ В.А., ХВОРОСТОВ П.В. Анализ метрических свойств графов //Методы обнаружения закономерностей с помощью ЭВМ. - Новосибирск, 1981. - Вып. 91: Вычислительные системы. - С. 3-20.

22. BALABAN A.T. Numerical modelling of chemical structures: local graph invariants and topological indices //Graph Theory and Topology in Chemistry (Studies in physical and theoretical chemistry, vol. 51) /Eds. King R.B. and Rouvray D.H. - Elsevier, 1987. - P. 159-176.

23. RANDIĆ M. Characterization of atoms, molecules and classes of molecules based on path enumeration //Comm. Math. Chem. (MATCH). - 1979. N. 7. -P. 5-64.

24. RANDIĆ M., WILKINS C.L. Graph theoretical approach to recognition of structural similarity in molecules //J.Chem. Inf. Comput. Sci. - 1979. - Vol. 19, N. 1. -P. 31-36.

25. РАНДИЧ М., КРАУС Дж., ДЗОНОВА-ДЖЕРМАН-БЛАЗИЧ Б. Упорядочивание графов как подход к исследованиям корреляций структура - активность. //Химические приложения топологии и теории графов. - М.: Мир, 1987. - С. 222-235.

26. MEKENYAN O., BONCHEV D., TRINAJSTIĆ N. Topological rules for spirocompounds //Comm. Math. Chem. (MATCH). - 1979. - N. 6. - P. 93-115.

27. BONCHEV D., MEKENYAN O., TRINAJSTIĆ N. Topological characterization of cyclic structures //Int. J. Quantum Chem. - 1980. - Vol. 17. - P. 845-893.

28. MEKENYAN O., BONCHEV D., TRINAJSTIĆ N. Algebraic characterization of bridged polycyclic compounds //Int. J. Quantum Chem. - 1981. - Vol. 19. - P. 929-955.

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

30. MEKENYAN O., BONCHEV D. Structural complexity and molecular properties of cyclic systems with acyclic branches // Croat. Chem. Acta. - 1983. - Vol. 56, N 2. -P. 237-261.

31. СКОРОБОГАТОВ В.А., МЖЕЛЬСКАЯ Е.В., МЕЙРМАНОВА Н.М.Изучение метрических характеристик ката-конденсированных поли - бензолов //Алгоритмический анализ графов и его применения.-Новосибирск, 1988. - Вып. 127: Вычислительные системы.-С. 40-91.

32. POLANSKY O.E., BONCHEV D. The Wiener number of graphs. I. General theory and changes due to some graph operations // Comm. Math. Chem. (MATCH). - 1986. - N 21. - P.133-186.

33. GUTMAN I., POLANSKY O.E. Wiener numbers of polyacens and related benzenoid molecules //Comm. Math. Chem.(MATCH). - 1986. - N 20. -P.115-123.

34. ДОБРЫНИН А.А. Дистанция графов ката-конденсированных полициклических систем при их преобразованиях //Алгоритмический анализ графов и его применения. - Новосибирск, 1988. - Вып. 127: Вычислительные системы. -С. 3-39.

35. POLANSKY O.E., BONCHEV D. Theory of the Wiener number of graphs. I.Transfer graphs and some of their metric properties //Comm. Math. Chem. (MATCH). - 1990. - N 25. -P. 3-39.

36. СКОРОБОГАТОВ В.А., ДОБРЫНИН А.А. Влияние структурных преобразований графа на значение его дистанции //Анализ данных в экспертных системах. - Новосибирск, 1986. -Вып. 117: Вычислительные системы. -С. 103-113.

37. GUTMAN I., MARKOVIĆ S., LUKOVIĆ U., RADIVOJEVIĆ V., RANČIĆ S. On Wiener numbers of benzenoid hydrocarbons //Zbornik Radova Prirodno - matematičkog faculteta u Kragujevcu. - 1987. - Vol. 8. - P.15-34.

38. ДОБРЫНИН А.А. Дистанция молекулярных графов полициклических соединений //Анализ данных в экспертных системах. - Новосибирск, 1986. - Вып. 117: Вычислительные системы. -С.114-122.

39. ДОБРЫНИН А.А. Распределения значений дистанции графов неразветвленных гексагональных систем //Математические иссле - дования в химической информатике. - Новосибирск,1990.-Вып.136: Вычислительные системы. - С. 61-141.

40. TRATCH S.S., STANKEVITCH M.I., ZEFIROV N.S. Combinatorial models and algorithms in chemistry. The expanded Wiener number - a novel topological index // J.Comput.Chem. - 1990. - Vol.11, № 8.- P.899-908.

41. ENTRINGER R.C., JACKSON D.E., SNYDER D.A. Distance in graphs //Czech.math.J.- 1976. - Vol.26, № 2.- P.283-296.

42. PLESNIK J. On the sum of all distances in a graph or digraph //J.Graph Theory. - 1984.- Vol. 8, № 1.-P.1-21.

43. BRIGHAM R.C., DUTTON R.D. A compilation on relations between graph invariants //Networks.- 1985.- Vol.15.-P.73-107.

44. BEHZAD M., SIMPSON J.E. Essentric sequences and essentric sets in graphs //Discrete Math.- 1976.- Vol.16.- P.187-193.

45. SLATER P.J. Centers to centroids in graphs // J.Graph Theory.- 1978.- Vol.2.- P.209-222.

46. HOWORKA E. A characterization of distance - hereditary graphs //Quart.J.Math.- 1977.- Vol.28,№ 112.-P.417-420.

47. BLOOM G.S., KENNEDY J.W., QUINTAS L.V. Some problems concerning distance and path degree sequences // Graph Theory: Proc./ Conference in Łagów, Poland, feb.1981. - Berlin a.o.: Springer, 1983.- P.179-190.- (Lecture Notes in Mathematics, 1018).

48. КАЦ А.О. Исследование системы векторов расстояний графа // Латв.мат.ежегодник.- Рига,1976.- Т.20.- С.170-179.

49. СКОРОБОГАТОВ В.А. Относительные разбиения и слои графов // Вопросы обработки информации при проектировании систем. - Новосибирск, 1977.- Вып.69: Вычислительные системы.- С.3-10.

50. QUINTAS L.V., SLATER P.J. Pairs of non-isomorphic graphs having the same path layer matrix // Comm. Math. Chem. (MATCH).- 1981.- № 12.- P.75-86.

51. SLATER P.J. Counterexamples to Randic's conjecture on distance degree sequences // J.Graph Theory.- 1982.- Vol.6,№1. - P.89-91.

52. DOBRYNIN A.A., Regular graphs having the same path layer matrix // J.Graph Theory.- 1990.- Vol.14,№ 2.-P.141-148.

53. ЗАГОРУЙКО Н.Г., СКОРОБОГАТОВ В.А., ХВОРОСТОВ П.В. Вопросы анализа и распознавания молекулярных структур на основе

общих фрагментов // Алгоритмы анализа структурной информации.- Новосибирск,1984.- Вып.103: Вычислительные системы.- С.26-50.

54. 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.92-129.

55. ЗАГОРУЙКО Н.Г., ЁЛКИНА В.Н., ЛБОВ Г.С. Алгоритмы обнаружения эмпирических закономерностей.- Новосибирск: Наука, 1985.- С.46-54.

56. Программный комплекс МЕТАХИМ / А.А.Добрынин,Е.В.Константинова, Л.И.Макаров и др.// Математические исследования в химической информатике.- Новосибирск, 1990.- Вып.136: Вычислительные системы.- С.3-15.

Поступила в ред.-изд.отд.

1 августа 1991 года

Значения метрических инвариантов некоторых графов

1. В <u>полном графе</u> $K_p$ дистанция вершины равна $D(v) = p-1$ и $D(K_p) = p(p-1)^2$, цепные дистанции

$$D_\tau(v) = (p-1)!\sum_{r=1}^{p-1} \frac{r}{(p-r-1)!}$$

и

$$D_\tau(K_p) = p(p-1)!\sum_{r=1}^{p-1} \frac{r}{(p-r-1)!},$$

эксцентриситеты $e(v) = 1$ и $e(K_p) = p$, цепные эксцентриситеты $e_\tau(v) = p-1$ и $e_\tau(K_p) = p(p-1)$.

2. В <u>звезде</u> $K_{1,p-1}$ для центральной вершины $D(v) = p - 1$, для других вершин $D(u) = 2p-3$ и $D(K_{1,p-1}) = 2(p-1)^2$, эксцентриситеты $e(v) = 1$, $e(u) = 2$ и $e(K_{1,p-1}) = 2p-1$. Значения цепных инвариантов совпадают с указанными.

3. <u>В полном графе без ребра</u> $K_p-x$ для вершины $v$ ребра $x$ $D(v) = p$ и для остальных вершин $D(u) = p-1$, для графа $D(K_p-x) = p^2-p+2$. Эксцентриситеты $e(v) = 2$, $e(u) = 1$ и $e(K_p-x) = p+2$. Цепные эксцентриситеты $e_\tau(v) = e_\tau(u) = p-1$ и $e_\tau(K_p-x) = p(p-1)$ для $p > 3$.

4. В <u>полном двудольном графе</u> $K_{p_1,p_2}$ дистанция вершин из доли порядка $p_1$ равна $D(v) = p_2+2p_1-2$, а для вершин из доли порядка $p_2$ есть $D(u) = p_1+2p_2-2$ и $D(K_{p_1,p_2}) = 2(p_1p_2+p_1(p_1-1)+p_2(p_2-1))$. Эксцентриситеты $e(v) = e(u) = 2$ и $e(K_{p_1,p_2}) = 2(p_1+p_2)$.

цепные эксцентриситеты при $p_1 > p_2$ равны $e_\tau(v) = 2p_2$, $e_\tau(u) = 2p_2-1$ и $e_\tau(K_{p_1,p_2}) = p_2(2p_1 + 2p_2 - 1)$.

5. В <u>простой цепи</u> $P = (v_1, v_2, \ldots, v_p)$ порядка $p$ дистанция вершины $v_i$ равна $D(v_i) = \frac{1}{2}p(p+1) - i(p+1-i)$, $D(v_1) = \frac{1}{2}p(p-1)$ и $D(v_{p/2}) = \frac{1}{4}(p^2-1)$ для нечетного $p$. Дистанция простой цепи $D(P) = \frac{1}{3}p(p^2-1)$. Эксцентриситет вершины $v_i$ равен $e(v_i) = p-i$, $i = 1, 2, \ldots, \lfloor p/2 \rfloor$. Эксцентриситет простой цепи дается выражением

$$e(P) = \begin{cases} \frac{1}{4}(p-1)(3p+1), & p \text{ нечетно,} \\ \frac{1}{4}p(3p-2), & p \text{ четно.} \end{cases}$$

6. Для <u>простого цикла</u> $C_p$ дистанция вершин равна

$$D(v) = \begin{cases} \frac{1}{4}p^2, & p \text{ четно,} \\ \frac{1}{4}(p^2-1), & p \text{ нечетно,} \end{cases}$$

и

$$D(C_p) = \begin{cases} \frac{1}{4}p^3, & p \text{ четно,} \\ \frac{1}{4}p(p^2-1), & p \text{ нечетно.} \end{cases}$$

Эксцентриситеты вершин $e(v) = p/2$ при четном $p$ и $e(v) = (p-1)/2$ при нечетном $p$. Соответственно $e(C_p) = p^2/2$ и $e(C_p) = p(p-1)/2$. Цепная дистанция вершин простого цикла есть $D_\tau(v) = p(p-1)$ и $D_\tau(C_p) =$

$= 2p^2(p-1)$. Цепные эксцентриситеты вершин $e_\tau(v) = p-1$ и $e_\tau(C_p) = p(p-1)$.

7. Для колеса $W_p$ дистанция центральной вершины $D(v) = p-1$, для других вершин $D(u) = 2p-5$ и $D(W_p) = 2(p-1)(p-2)$. Эксцентриситеты $e(v) = 1$ $e(u) = 2$ и $e(W_p) = 2p-1$.

8. Для призмы $Pr_n$ с $n$-угольником в основании и порядком $p$

$$D(Pr_n) = \begin{cases} n(n^2+2n-1), & p = 2n+1, \\ n^2(n+2), & p = 2n. \end{cases}$$

Эксцентриситет вершины $e(v) = \frac{1}{2}(n+2)$ при четном $n$ и $e(v) = \frac{1}{2}(n+1)$ при четном $n$. Цепные эксцентриситеты призмы $e_\tau(v) = p-1$ и $e_\tau(Pr_n) = p(p-1)$.

9. При операции соединения произвольного графа $G$ радиуса $r(G) > 1$ с графом $K_1$ дистанция центральной вершины в полученном графе $K_1 \cdot G$ порядка $p$ равна $D(v) = p-1$, для других вершин $K_1 \cdot G$ имеем $D(u) = 2p - \deg(u) - 2$ и $D(K_1 \cdot G) = 2(p^2-p-q) = 2(p_G^2 - q_G)$. Эксцентриситеты $e(v) = 1$, $e(u) = 2$ и $e(K_1 \cdot G) = 2p-1 = 2p_G+1$.

10. Для вершин изображенного на рис.1 графа диаметра $d$ и порядка $p$ выполняется $D(v_i) = p(d-i+1) - \frac{1}{2}d(d+1) + i(i-1)$ для $i = 1, 2, \ldots, d$ и $D(v_i) = \frac{1}{2}d(d-3) + 2p-2$ для $i = d+1, d+2, \ldots, p$. Дистанция графа равна $D(G) = pd(d-3) - \frac{1}{3}d(2d^2-3d-5) + 2p(p-1)$. Если порядок графа есть полный квадрат, то для диаметра $d = \sqrt{p}$

имеем $D(v_1) = \frac{1}{2}\sqrt{p}(2p-\sqrt{p}-1)$ и $D(G) = \frac{1}{3}\sqrt{p}(9p\sqrt{p}-11p-3\cdot\sqrt{p}+5)$.

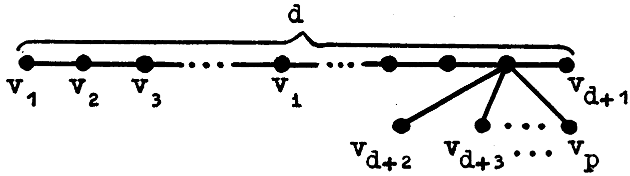



Рис. 1

11. В графах полициклических ката-конденсированных систем (два кольца имеют только единственное общее ребро) с кольцами произвольного размера эксцентриситеты вершин равны $e(v) = p-1$ и для графа $e(G) = p(p-1)$.

12. Для графов неразветвленных гексагональных систем на рис.2, состоящих из $h$ соединенных по ребру шестиугольников, дистанции имеют следующие экстремальные значения:

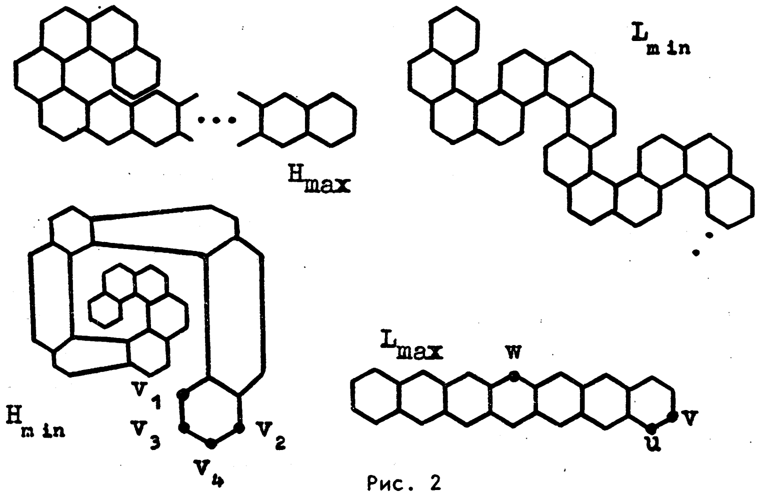



Рис. 2

а) графы, вложимые в правильную гексагональную решетку на плоскости с наибольшим значением дистанции

$$D(L_{max}) = \frac{2}{3}(16h^3 + 36h^2 + 26h + 3).$$

Наибольшие дистанции вершин $D(v) = 4h^2 + 4h + 1$ и $D(u) = 4h^2 + 5$, наименьшая дистанция вершин $D(w) = 2h^2 + 4h + 1$ при четном $h$ и $D(w) = 2h^2 + 4h + 3$ при нечетном $h \geq 3$. Диаметр и центр графа $r(L_{max}) = h+1$ и $d(L_{max}) = 2h+1$, эксцентриситет графа $e(L_{max}) = 2(h+1)(3h+2)$ при четном $h \geq 2$ и $e(L_{max}) = 2(h+1)(3h+1)$ при нечетном $h \geq 3$;

б) графы, вложимые в правильную гексагональную решетку на плоскости с наименьшим значением дистанции

$$D(L_{min}) = \frac{2}{9}(32h^3 + 168h^2 + \varphi(h)),$$

где

$$\varphi(h) = \begin{cases} -6h + 81, & h = 3m, \ m = 1,2,3,\ldots, \\ -6h + 49, & h = 3m+1, \ m = 0,1,2,\ldots, \\ -54h + 161, & h = 3m+2, \ m = 0,1,2,\ldots; \end{cases}$$

в) графы, не вложимые в правильную гексагональную решетку на плоскости с наименьшим значением дистанции

$$D(H_{min}) = \frac{2}{3}(8h^3 + 72h^2 - 26h + 27).$$

Дистанции вершин степени 2 в концевом кольце графа $D(v_1) = 2h^2 + 6h + 1$, $D(v_2) = 2h^2 + 10h - 7$, $D(v_3) = 2h^2 + 10h - 3$ и $D(v_4) = 2h^2 + 14h - 11$ при $h \geq 2$. Диаметр и радиус графа $d(H_{min}) = h+4$, $r(H_{min}) = (h+4)/2$ при четном $h \geq 2$ и $r(H_{min}) = (h+3)/2$ при нечетном $h \geq 3$;

г) графы, не вложимые в правильную гексагональную решетку на плоскости с наибольшим значением дистанции

$$D(H_{max}) = \frac{2}{3}(16h^3 + 36h^2 - 358h + 1587) + \varphi(h),$$

где

$$\varphi(h) = \begin{cases} 8, & h = 8, \\ 0 - & \text{в противном случае.} \end{cases}$$

Диаметр и радиус графа $d(H_{max}) = 2(h-1)$ и $r(H_{max}) = h-1$ при $h \geq 6$.

13. Для графов неразветвленных пентагональных систем, состоящих из $h$ соединенных по ребру пятиугольников (рис.3), имеем:

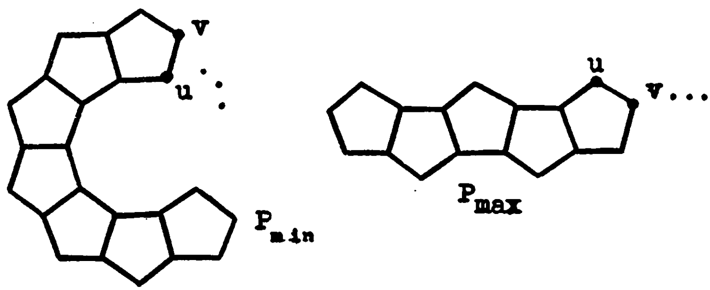



Рис. 3

а) для графа $P_{min}$ с наименьшей дистанцией выполняется

$$D(P_{min}) = 3h(h+1)(h+6) - 24h + 14.$$

Дистанции вершин $D(u) = \frac{1}{2}(3h^2 + 7h + 2)$ и $D(v) = \frac{1}{2}(3h^2 + 13h - 6)$ ;

б) для графа $P_{max}$ с наибольшей дистанцией выполняется

$$D(P_{max}) = \frac{1}{4}(18h^3 + 45h^2 + 58h + \varphi(h)),$$

где

$$\varphi(h) = \begin{cases} -1, & h \text{ нечетно,} \\ 0, & h \text{ четно.} \end{cases}$$

Дистанции вершин $D(u) = \frac{1}{4}(9h^2 + 6h + c)$, где $c = 8$ при $h$ четном и $c = 9$ при $h$ нечетном, и $D(v) = \frac{1}{4}(9h^2 + 12h + c)$, где $c = 4$ при $h$ четном и $c = 3$ при $h$ нечетном.