МАТЕМАТИЧЕСКИЕ ИССЛЕДОВАНИЯ В ХИМИЧЕСКОЙ ИНФОРМАТИКЕ
(Вычислительные системы)
1990 год
Выпуск 136

УДК 519.17:547.64

# О ЧУВСТВИТЕЛЬНОСТИ ТОПОЛОГИЧЕСКИХ ИНДЕКСОВ ПОЛИЦИКЛИЧЕСКИХ ГРАФОВ

Е.В.Константинова, А.А.Палеев

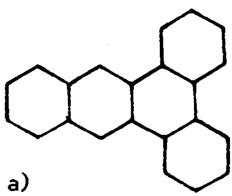
## Введение

Одной из классических задач теории графов является установление изоморфизма двух графов [1]. Для однозначной характеризации молекулярных графов в последнее время используются топологические индексы [2,3]. Оценкой способности распознавать неизоморфные графы данным топологическим индексом $I$ служит его чувствительность [4-6], характеризующая возможность правильной идентификации графов. В настоящей работе рассматривается чувствительность топологических индексов полициклических графов, являющихся подграфами правильных четырехугольной $R^4$ и шестиугольной $R^6$ решеток [7-9]. Известно [10-12], что топологические индексы, как правило, терпят вырождение на этих графах. В частности, в [12] изучалась возможность идентификации полициклических подграфов $R^6$ метрическими индексами. Ни один из рассмотренных десяти индексов не позволяет это сделать однозначно. Среди рассмотренных в данной работе 14 топологических индексов выделено 4 индекса, чувствительность которых позволяет использовать их в качестве фильтров в задаче установления изоморфизма графов.

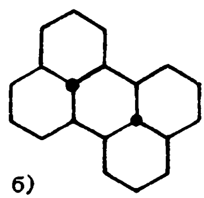
## 1. Определение полициклических графов

Пусть $G(V,E)$ есть конечный неориентированный связный граф без петель и кратных ребер с множеством вершин $V(G)$, $|V(G)| = p$. Граф G называется плоским топологическим графом, если существует его изображение на плоскости, не содержащее пересечений ребер. Гранью G называется область плоскости, ограниченная ребрами и не содержащая внутри себя ни вершин, ни ребер.

Рассмотрим бесконечный плоский топологический граф $R^6$, все грани которого есть правильные шестиугольники, а вершины имеют степень три. Пусть G есть подграф $R^6$, границей которого является простой цикл. Граф G называется ката-конденсированным бензоидным графом*) [7-9], если ни одна из его вершин не является



а)



б)

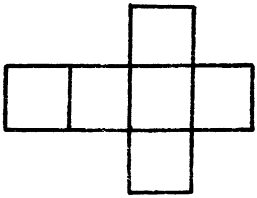
Рис. 1

внутренней (рис.1а), граф G называется пери-конденсированным бензоидным графом, если он содержит хотя бы одну внутреннюю вершину (рис.1б). Множество всех ката-конденсированных и пери-конденсированных графов называют К6-графами.

Аналогичным образом определяются [8,9] подграфы плоского топологического графа $R^4$, все грани которого есть квадраты, а вершины имеют степень четыре (рис.2). Множество всех таких подграфов $R^4$ называют К4-графами. В дальнейшем множество К4- и

---

*) Название графов взято из химии полициклических углеводородов, структурным формулам которых данные графы соответствуют.



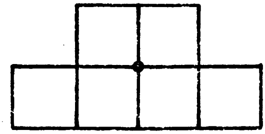


Рис. 2

К6-графов будем называть просто полициклическими графами. Изображение полициклических графов с числом вершин $1 \leq h \leq 9$ для К6-графов и $1 \leq h \leq 7$ для К4-графов можно найти в [12*), 13] и [8] соответственно.

## 2. Топологические индексы графов

Всякому графу можно поставить в соответствие набор инвариантов, т.е. совокупность чисел, не зависящих от способа нумерации вершин графа. Инварианты молекулярных графов - графов, представляющих органические молекулы, - называют топологическими индексами [2,3].

Первый топологический индекс был предложен в 1947 г. Винером [14] и носит имя его создателя. Индекс Винера определяется формулой

$$W(G) = \frac{1}{2} \sum_{i,j=1}^{p} d_{ij} , \qquad (1)$$

где $d_{ij}$ есть расстояние между вершинами $i$ и $j$, равное числу ребер в кратчайшей $(i,j)$-цепи.

---

*) В [12] даны изображения только ката-конденсированных бензоидных графов с числом циклов $3 \leq h \leq 8$.

В настоящее время одним из важных вопросов приложения теории графов является исследование теоретико-информационных индексов графов [2,3,10]. Интерес в изучении топологических индексов, основанных на теории информации, объясняется, в частности, тем, что информационные индексы имеют обычно большую дискриминирующую способность по сравнению с соответствующими топологическими индексами [2].

Общим при построении информационных индексов является следующий известный принцип.

Пусть X есть некоторое множество, состоящее из $n$ элементов. Предположим, что по некоторому критерию эквивалентности элементы множества разбиваются на $N$ классов эквивалентности $X_i$ так, что $n = \sum_{i=1}^{N} n_i$, где $n_i$ - число элементов подмножества $X_i$. Тогда величина $p_i = n_i/n$ есть вероятность попадания одного элемента в $i$-е подмножество и для количественной оценки информации, содержащейся в одном элементе множества, можно использовать энтропию распределения вероятностей элементов этого множества, которая определяется формулой Шеннона:

$$H = -\sum_{i=1}^{N} p_i \log_2 p_i. \qquad (2)$$

При построении информационных индексов используются различные типы матриц графов [3], строятся информационные аналоги известных топологических индексов. В частности, сконструирован информационный аналог индекса Винера [15]:

$$I_D^W = -\sum_{i=1}^{N} \frac{n_i \cdot i}{W} \cdot \log_2 \frac{n_i \cdot i}{W}, \qquad (3)$$

где $n_i$ - число пар вершин, находящихся на расстоянии $i$.

Известные информационные индексы, построенные на основе матриц графов, отражают меру сложности графа в целом. Однако они могут использоваться и для характеризации отдельных ее элементов, в частности, вершин молекулярного графа. Ниже на примере двух матриц: матрицы расстояний $D = D(G)$ и матрицы слоев $\lambda = \lambda(G)$ показана возможность построения таких индексов.

Матрицей расстояний $D = \| d_{ij} \|$ графа $G$ называется $(p \times p)$-матрица, элементом $d_{ij}$ которой является расстояние между вершинами $i$ и $j$.

Матрицей слоев $\lambda = \| \lambda_{ij} \|$ графа $G$ называется $(p \times d(G))$-матрица, где $d(G)$ - диаметр графа, в которой $\lambda_{ij}$ равно числу вершин, находящихся на расстоянии $j$ от вершины $i$, $j = \overline{0, d(G)}$. Множество вершин, находящихся на расстоянии $j$ от данной вершины, называют $j$-слоем вершины.

Пусть множество $X$ вершины $v_i$ определяют ненулевые элементы соответствующей строки матрицы слоев. Из свойств матрицы следует, что с учетом вершин, находящихся на расстоянии 0 от данной, $|X| = p$. Тогда $p_{ij} = \frac{\lambda_{ij}}{p}$ есть вероятность случайно выбранной вершины попасть в $j$-слой вершины $i$, $j = \overline{0, e(v_i)}$, и слоевой информационный индекс вершины $v_i$ определяется формулой

$$H_\lambda(v_i) = -\sum_{j=0}^{e(v_i)} \frac{\lambda_{ij}}{p} \cdot \log_2 \frac{\lambda_{ij}}{p} . \tag{4}$$

Аналогичным образом определяется информационный индекс вершины по матрице расстояний. Вероятность случайно выбранной вершины находиться на расстоянии $d_{ij}$ от вершины $i$ есть величина $p_{ij} = \frac{d_{ij}}{d(v_i)}$, где $d(v_i) = \sum_{j=1}^{p} d_{ij}$, и дистанционный информационный индекс вершины $v_i$ определяется по формуле:

$$H_d(v_i) = -\sum_{j=1}^{p} \frac{d_{ij}}{d(v_i)} \cdot \log_2 \frac{d_{ij}}{d(v_i)} . \tag{5}$$

На основе представленных выше информационных индексов вершин можно получить также интегральные информационные индексы:

дистанционный индекс вершин графа $H_d = \sum_{i=1}^{p} H_d(v_i)$, (6)

слоевой индекс вершин графа $H_\lambda = \sum_{i=1}^{p} H_\lambda(v_i)$. (7)

Отметим, что в отличие от матрицы расстояний матрица слоев, учитывая метрическое распределение вершин относительно каждой вершины графа, не содержит самих значений расстояний. И в этом смысле является менее информативной по сравнению с матрицей расстояний, что отражается на значениях соответствующих информационных индексов.

В следующем разделе приведены результаты оценки чувствительности предложенных информационных индексов и некоторых наиболее известных топологических индексов [2,3] на классе полициклических графов.

## 3. Оценка чувствительности топологических индексов полициклических графов

Чувствительность топологического индекса I есть мера его способности распознавать неизоморфные графы [6]. Теоретическая оценка чувствительности I на множестве всех графов достаточно сложна, поэтому на практике используется оценка S чувствительности I на фиксированном множестве неизоморфных графов M:

$$S = \frac{N-N_I}{N}, \qquad (8)$$

где $N_I$ есть число вырождений I на множестве M, $N = |M|$.

Исследовалась чувствительность информационных индексов $I_D^W$, $H_d$, $H_\lambda$, описанных в предыдущем разделе, а также чувстви -

тельность 11 топологических индексов, ссылки на которые даны в таблице.

В качестве множества неизоморфных графов M были рассмотрены следующие множества полициклических графов:

$M_1$: все К6-графы с числом циклов $h = \overline{1,7}$, а также катаконденсированные графы с числом циклов $h = 8$. Мощность данного множества графов $N_1 = 857$;

$M_2$: все К4-графы с числом циклов $h = \overline{1,7}$, $N_2 = 163$;

$M_3$: объединенное множество $M_3 = M_1 \cup M_2$, $N_3 = 1020$.

Результаты вычислений, полученные на IBM PC/AT, представлены в таблице. Точность вычислений данной ПЭВМ - 16 значащих цифр. Наиболее чувствительными на данных классах графов являются информационные индексы и индекс Балабана, наименее чувствительными - целочисленные индексы Платта, Гордона-Скантлбери, загребской группы и полной смежности. Причиной этого является то, что индексы первой группы не ограничены целочисленными величинами и получаются при суммировании большего числа различных величин, чем в случае целочисленных топологических индексов. Индекс Балабана и информационные индексы $H_d$ и $I_D^W$ полностью идентифицируют класс К6-графов, однако на классе К4-графов эти индексы вырождаются. Индекс Балабана, в частности, вырождается на графах, изображенных на рис.3; его значение на этих графах равно 0.1516360807197543.

$G_1$ $G_2$

Рис. 3

Число значащих цифр $k$, необходимое для различия значений индекса $I$ на рассмотренных множествах графов, для различных

Т а б л и ц а

Оценка чувствительности топологических индексов на множестве полициклических графов

| И н д е к с ы | Множество графов | | | | | |
|---|---|---|---|---|---|---|
| | $M_1$: К6-графы $N_1 = 857$ | | $M_2$: К4-графы $N_2 = 163$ | | $M_3$: К6- и К4-графы $N_3 = 1020$ | |
| | $N_I$ | S | $N_I$ | S | $N_I$ | S |
| Индекс Балабана J(G) [2] | 0 | 1 | 2 | 0.988 | 2 | 0.998 |
| $H_d$ | 0 | 1 | 4 | 0.975 | 4 | 0.996 |
| $I_D^W$ | 0 | 1 | 6 | 0.963 | 8 | 0.992 |
| $H_\lambda$ | 2 | 0.998 | 10 | 0.939 | 14 | 0.986 |
| Среднее дистанционное отклонение графа $\Delta D(G)$ [10] | 91 | 0.894 | 89 | 0.454 | 190 | 0.823 |
| Индекс расстояния между вершинами VDI(G) [2] | 307 | 0.642 | 93 | 0.429 | 400 | 0.608 |
| Индекс среднеквадратичных расстояний $D^{(2)}(G)$ [2] | 339 | 0.604 | 100 | 0.387 | 441 | 0.568 |
| Индекс Винера W(G) [2] | 634 | 0.260 | 122 | 0.252 | 758 | 0.257 |
| Индекс связности Рандича $\chi(G)$ [2] | 810 | 0.055 | 51 | 0.687 | 861 | 0.156 |
| Эксцентриситет графа e(G) [10] | 723 | 0.156 | 137 | 0.160 | 867 | 0.150 |
| Индекс загребской группы $M_1(G)$ [2] | 849 | 0.009 | 154 | 0.055 | 1005 | 0.015 |
| Индекс Платта F(G) [2] | 849 | 0.009 | 155 | 0.049 | 1009 | 0.011 |
| Индекс Гордона-Скантлбери Y(G) [2] | 849 | 0.009 | 155 | 0.049 | 1009 | 0.011 |
| Индекс полной смежности A'(G) [2] | 850 | 0.008 | 158 | 0.031 | 1013 | 0.007 |

индексов различно. Так, для различия почти всех значений информационно-дистанционного индекса $H_d$ достаточно 4 значащих цифры, за исключением одной пары графов, для которых расхождение в значении индекса наблюдается в пятой цифре. Требуемая точность вычислений индекса Балабана и информационного индекса Винера $I_D^W$ соответственно равны $k = 8$ и $k = 7$. Значение информационно-слоевого индекса $H_\lambda$ на двух парах К4-графов различаются в пятой и одиннадцатой цифрах, при этом для всех остальных К6- и К4-графов расхождение в значении индекса наблюдается в четвертой цифре.

Отметим также, что чувствительность почти всех индексов на классе К6-графов выше, чем на классе К4-графов, за исключением индекса Рандича. Последний определяется на основе степеней вершин в графе, и поскольку степени вершин К4-графов представлены тремя значениями (2,3,4), а степени вершин К6-графов только двумя (2,3), то это объясняет достаточно высокую чувствительность индекса Рандича на К4-графах по сравнению с К6-графами.

В целом по результатам проведенного анализа чувствительности топологических индексов на множестве полициклических графов предпочтение в задачах идентификации графов следует, видимо, отдавать информационным, а также нецелочисленным топологическим индексам, учитывающим более полно структурные особенности графов.

## Заключение

В работе исследована чувствительность 14 топологических индексов на классе полициклических графов. Максимальная мощность рассмотренных множеств графов - 1020. Выделено 4 наиболее чувствительных индекса, среди которых 3 являются информационными. Чувствительность данных индексов меняется в пределах от 0.939 до 1.

Авторы выражают признательность Некрасову Ю.С. и Тепферу Э.Э., совместная работа с которыми послужила развитию нового направления в наших исследованиях, а именно - разработки и изучения теоретико-информационных индексов.

## Литература

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

2. РУВРЭ Д. Следует ли заниматься разработкой топологических индексов? //Химические приложения топологии и теории графов: Пер. с англ./Под ред. Р.Кинга. - М.: Мир, 1987. - С. 183-205.

3. СТАНКЕВИЧ М.И., СТАНКЕВИЧ И.В., ЗЕФИРОВ Н.С. Топологические индексы в органической химии //Успехи химии. - 1988.- Т. 57, вып. 3. -С. 337-366.

4. BONCHEV D., MEKENYAN O., TRINAJSTIČ N. Isomer descrimination by topological information approach //J.Comp. Chem. - 1981. - Vol. 2, N 2. - P. 127-148.

5. ДРБОГЛАВ В.В., ГОЛЕНДЕР В.Е. Сравнение эффективности топологических индексов в задаче идентификации химических структур //Тез.докл. 7 Всесоюзной конференции по использованию вычислительных машин в химических исследованиях и спектроскопии молекул. - Рига, 1986. - С. 219-220.

6. ДРБОГЛАВ В.В. Инварианты графов и их использование для обработки структурной информации: Автореф.дис... канд.техн.наук: 05.13.16. - Новосибирск, 1987. - 232 с.

7. МЖЕЛЬСКАЯ Е.В. Графы полициклических соединений //Вопросы алгоритмического анализа структурной информации. - Новосибирск, 1987. - Вып. 119: Вычислительные системы. -С. 71-90.

8. МЖЕЛЬСКАЯ Е.В. Подграфы правильной четырехугольной решетки $R^4$ //Алгоритмический анализ графов и его применения. - Новосибирск, 1988. - Вып. 127: Вычислительные системы. -С.143-151.

9. МЖЕЛЬСКАЯ Е.В., СКОРОБОГАТОВ В.А. Применение теории графов в химии полициклических бензоидных углеводородов. - Новосибирск, 1987. - 33 с. (Препринт/АН СССР. Сиб.отд-ние.Ин-т математики; № 35).

10. БОНЧЕВ Д.Г. Характеризация химических структур с помощью теории информации и теории графов: Автор.дис.д-ра. хим. наук. - Бургас, 1983. - 48 с.

11. GUTMAN I., MARKOVIĆ S., LUKOVIĆ U., RADIVOJEVIĆ V., RANČIĆ S. On Wiener numbers of benzenoid hydrocarbons //Zbor - nic Radova Prirodno-matematickog fakulteta u Kraqujevcu. -1987. - N 8. -P. 15-34.

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

13. KNOP J.V., MÜLLER W.R., SZYMANSKI K., TRINAJSTIČ N. Computer Generation of Certain Classes of Molecules. - Zagreb: SKTH-Kemija u industriji, 1985. - 166 p.

14. WIENER H. Structural determination of paraffin boi - ling points //J.Am.Chem.Soc. - 1947. -Vol. 69. -P. 17-20.

15. BONCHEV D., TRINAJSTIČ N. Information Theory, Distance Matrix and Molecular Branching //J.Chem.Phys. - 1987. -Vol. 67. - P. 4517-4533.

Поступила в ред.-изд.отд.

31 июля 1990 года