ЖУРНАЛ СТРУКТУРНОЙ ХИМИИ

Том 40, № 2 *Март – апрель* *1999*

УДК 547.1:51–7

А.А. ДОБРЫНИН, И. ГУТМАН, В.Н. ПИОТТУХ-ПЕЛЕЦКИЙ

ГИПЕРИНДЕКС ВИНЕРА ДЛЯ АЦИКЛИЧЕСКИХ СТРУКТУР

Исследуется дискриминирующая сила (т.е. способность различать изомеры) нового топологического индекса — гипериндекса Винера на классе ациклических структур (деревьев), включающих молекулярные графы алканов. Для деревьев с количеством вершин от 4 до 26 получены численные данные о классах деревьев с совпадающими значениями индекса. Для максимальных по мощности классов приведены диаграммы соответствующих деревьев. Обсуждаются применения индекса для решения задач обработки структурной химической информации.

Топологические индексы графов привлекают внимание химиков как инструмент для компактного и эффективного описания структурных формул химических молекул, позволяющий также изучать и прогнозировать взаимосвязь строения и свойств органических соединений. В работе изучается недавно введенный инвариант графов, представляющий интерес для теоретической химии. В честь американского химика Г. Винера он называется гипериндексом Винера и обозначается WW. Этот инвариант предложен Рандичем несколько лет назад как обобщение понятия классического индекса Винера W [1]. Напомним, что W определяется как сумма расстояний между всеми парами вершин графа. Индекс Винера играет заметную роль в теоретической химии и изучается с 1947 г. Работа Рандича стимулировала большое количество новых исследований, посвященных WW [2 — 11].

Наиболее хорошо топологические индексы зарекомендовали себя при описании и прогнозировании свойств ациклических соединений — алканов и их галоидпроизводных. Особенностью этого класса соединений является возможность их представления специфическим видом графов — деревьями.

Исходным пунктом для введения понятия гипериндекса Винера послужило следующее свойство индекса Винера для деревьев [12, 13]. Пусть T есть дерево с $n(T)$ вершинами, которые занумерованы числами 1, 2,..., n. Для ребра e с концевыми вершинами u и v будем использовать запись $e = (u, v)$. Обозначим $n_u(e|T)$ число вершин дерева T, лежащих ближе к вершине u, чем к вершине v (вершина u также учитывается). Аналогично, пусть $n_v(e|T)$ есть количество вершин дерева T, лежащих ближе к вершине v, чем к вершине u. Эти величины также равны числу вершин в двух деревьях (компонентах связности), образующихся после удаления ребра (u, v) из исходного дерева. Очевидно, что для любого ребра e выполняется $n_u(e|T) + n_v(e|T) = n(T)$. Тогда индекс Винера для дерева T удовлетворяет следующему равенству [12]:

$$W(T) = \sum_e n_u(e|T) n_v(e|T) ,$$

где суммирование ведется по всем ребрам T, т.е. по всем парам смежных вершин дерева.

В дереве любые две вершины соединены единственной цепью. Заметим, что величина $n_u(e|T)$ равна числу таких вершин x дерева T, что цепь между x и u не содержит ребра e. Аналогично, $n_v(e|T)$ есть количество таких вершин y дерева T, для которых цепь между y и v не включает e. Такая интерпретация чисел $n_u(e|T)$ и $n_v(e|T)$ непосредственно переносится на случай, когда u и v являются произвольными вершинами дерева. Пусть p есть цепь, соединяющая вершины u и v. Тогда $n_u(p|T)$ определяется как число таких вершин x дерева T, для которых цепь между вершинами x и u не имеет ни одного общего ребра с цепью p. Аналогично, $n_v(p|T)$ равно количеству таких вершин y дерева T, для которых цепь между y и v не содержит общих ребер с цепью p. Числа $n_v(p|T)$ и $n_u(p|T)$ можно понимать как количества вершин в компонентах связности, получаемых после удаления из дерева цепи p и всех ветвей, присоединенных к внутренним вершинам p. Очевидно, что если u и v смежны, то $n_u(p|T) = n_u(e|T)$ и $n_v(p|T) = n_v(e|T)$.

Гиперииндекс Винера дерева T определяется с использованием введенных величин следующим образом [1]:

$$WW(T) = \sum_{p} n_u(p|T) n_v(p|T)\,, \tag{1}$$

где суммирование ведется по всем цепям дерева T, т.е. по всем парам вершин T. Если в качестве p рассматривать только двухвершинные цепи, то WW совпадает с W.

Как уже было упомянуто, индекс Винера есть сумма расстояний между всеми парами вершин графа:

$$W(G) = \sum_{x<y} d(x, y|G)\,,$$

где $d(x, y|G)$ обозначает расстояние между вершинами x и y в графе G, равное длине кратчайшей по числу ребер цепи, соединяющей x и y.

Недавно показано, что в случае деревьев выполняется следующее равенство [6]:

$$WW(T) = \frac{1}{2}\left[\sum_{x<y} d(x, y|T)^2 + \sum_{x<y} d(x, y|T)\right], \tag{2}$$

которое и было использовано для перенесения понятия гипериндекса Винера на случай произвольных связных графов. Заметим, что определить WW для произвольного графа равенством (1) затруднительно, так как пара вершин может быть соединена в графе несколькими кратчайшими цепями.

В работе изучаются свойства гипериндекса Винера для класса деревьев, т.е. ациклических связных графов без петель и кратных ребер. Основное внимание уделено классам неизоморфных деревьев с совпадающим значением WW.

ДИСКРИМИНИРУЮЩАЯ СИЛА ТОПОЛОГИЧЕСКИХ ИНДЕКСОВ

В химии структура молекулы представляется так называемым молекулярным графом. Вершины в таком графе соответствуют атомам, а ребра — химическим связям молекулы. Любой (численный) инвариант молекулярного графа, оказывающийся полезным в тех или иных задачах химии, называется топологическим индексом [13 — 15]. Большинство топологических индексов используется для моделирования различных физико-химических свойств органических соединений, в том числе имеющих значение для фармакологии. С этой точки зрения индекс Винера и гипериндекс Винера являются типичными топологическими индексами.

Топологические индексы, как правило, отражают структуру описываемой молекулы и часто рассматриваются как структурные дескрипторы. Ввиду этого желательно, чтобы молекулы с различающейся структурой имели различные значения топологического индекса. Это особенно важно для изомеров — молекул, содержащих одинаковое число атомов каждого вида, но отличающихся по структуре. Способность топологического индекса различать изомеры называется дискриминирующей силой индекса.

Ни один из известных топологических индексов не обладает абсолютной дискриминирующей силой, т.е. всегда существуют изомеры, на которых численные значения данного топологического индекса совпадают. Более того, почти всегда существует возможность найти семейства графов с зафиксированным числом вершин и ребер таких, что данный инвариант принимает одинаковое значение для всех членов этого семейства. При этом говорят, что инвариант вырождается на соответствующих графах, а само семейство называют классом вырождения инварианта. Одним из направлений химической теории графов является построение полезных (для химии) структурных дескрипторов, имеющих большую дискриминирующую силу.

Известно, что дискриминирующая сила индекса Винера довольно мала [16 — 18]. Недавно показано, что, по крайней мере, в случае деревьев дискриминирующая сила гипериндекса Винера также должна быть небольшой [11]. Это непосредственно следует из того факта, что среди n-вершинных деревьев индекс WW достигает своего минимума и максимума на звезде S_n и цепи P_n с n вершинами соответственно [11]. Именно,

$$WW(S_n) = \frac{1}{2}(n-1)(3n-4),$$

$$WW(P_n) = \frac{1}{24}n(n-1)(n+1)(n+2).$$

Отсюда следует, что при возрастании порядка деревьев количество чисел, принадлежащих целочисленному интервалу $[WW(S_n),\ WW(P_n)]$, становится много меньше, чем число n-вершинных деревьев. Это, в свою очередь, значит, что гипериндекс Винера имеет ограниченное применение как структурный дескриптор и его использование при изучении взаимосвязи структура — свойство должно осуществляться с известной осмотрительностью.

ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Для исследования дискриминирующей силы гипериндекса Винера были вычислены значения WW для деревьев с числом вершин $4 \le n \le 26$. На основе этих данных получены распределения числа деревьев по значениям WW для $4 \le n \le 13$ (полные численные данные можно получить у авторов).

Первые случаи вырождения гипериндекса Винера наблюдаются на деревьях с 9 вершинами (для индекса Винера — на деревьях с 7 вершинами). Диаграммы деревьев порядка $9 \le n \le 12$ из всех максимальных классов вырождения WW приводятся на рис. 1. Под деревьями указаны значения WW. Отметим, что существуют деревья с различающимся числом вершин и одинаковым значением WW. Так, деревья с $WW = 213$ встречаются при $n = 9, 10, 11$.

Количество классов вырождения с возрастанием числа вершин деревьев увеличивается очень быстро. Распределения мощности классов вырождения WW получены для деревьев порядка $9 \le p \le 17$. Рост количества и мощности классов вырождения в среднем и максимальных классов можно оценить по данным таблицы.

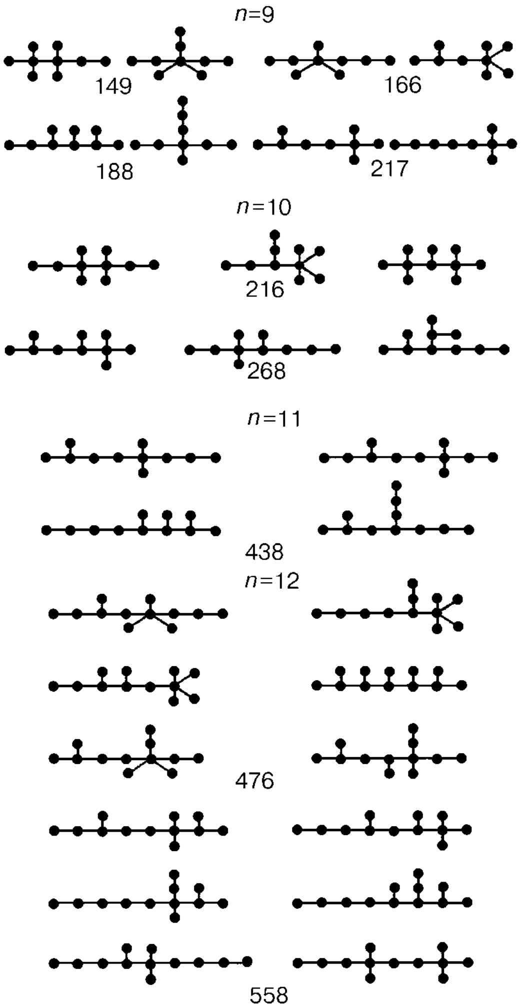

Рис. 1. Деревья с числом вершин $9 \le n \le 12$ из максимальных классов вырождения гипериндекса Винера

Она является расширенной версией аналогичной таблицы из [11]. Здесь $|E_n|$ есть длина целочисленного интервала $E_n = [WW(S_n), WW(P_n)]$, t_n равно количеству n-вершинных деревьев, *Max* есть размер максимального класса вырождения *WW*. Обозначим R_n множество тех значений из E_n, которые реализуются как значения гипериндекса Винера деревьев, т.е. $R_n = \{WW(T) | T \in \omega_n\}$, где ω_n есть множество всех деревьев на n вершинах. Тогда *Est* есть нижняя оценка мощности класса вырождения в среднем, $Est = t_n/|E_n|$. Величина *Avr* дает фактическую мощность класса вырождения в среднем ($Avr = t_n/|R_n|$). В последней колонке таблицы приводится количество классов вырождения *WW*, содержащих более одного дерева (*Ntr*).

Различие чисел в столбцах *Est* и *Avr* зависит от величины $|E_n \backslash R_n|$. Покажем, что существуют интервалы нереализуемых значений из E_n, имеющие линейную по n длину. Рассмотрим деревья на рис. 2. В [11] утверждается, что для произвольного дерева T_n ($n \ge 7$), отличного от деревьев на рис. 2, выполняется неравенство

$$WW(S_n) < WW(S_n') < WW(S_n'') < WW(T_n) < WW(P_n'') < WW(P_n') < WW(P_n).$$

Нетрудно подсчитать, что

$$WW(S_n') = \frac{1}{2}(3n^2 - n - 14),$$

$$WW(S_n'') = \frac{1}{2}(3n^2 + 5n - 44),$$

$$WW(P_n'') = \frac{1}{24}(n+2)(n^3 - 25n + 96),$$

$$WW(P_n') = \frac{1}{24}(n+2)(n^3 - 13n + 36).$$

Мощность классов вырождения для WW

n	$\mid E_n \mid$	t_n	*Est*	*Avr*	*Max*	*Ntr*
1	1	1	1,00	1,00	1	0
2	1	1	1,00	1,00	1	0
3	1	1	1,00	1,00	1	0
4	4	2	1,00	1,00	1	0
5	14	3	1,00	1,00	1	0
6	36	6	1,00	1,00	1	0
7	76	11	1,00	1,00	1	0
8	141	23	1,00	1,00	1	0
9	239	47	1,00	1,09	2	4
10	379	106	1,00	1,16	3	13
11	571	235	1,00	1,39	4	53
12	826	551	1,00	1,78	6	143
13	1156	1301	1,13	2,36	10	319
14	1574	3159	2,01	3,68	17	623
15	2094	7741	3,70	6,29	26	1001
16	2731	19320	7,07	11,30	50	1429
17	3501	48629	13,89	21,48	86	2003
18	4421	123967	28,02	40,76	160	2684
19	5509	317955	57,72	81,46	319	3517
20	6784	823065	121,32	166,24	657	4517
21	8266	2144505	259,44	334,97	1391	5737
22	9976	5623756	563,73	744,28	3184	7022
23	11936	14828074	1242,30	1591,85	6714	8749
24	14169	39299897	2773,65	3505,48	15510	10489
25	16699	104636890	6266,06	7791,86	35956	12737
26	19551	279793450	14310,95	17572,76	84827	15157

Полученные численные данные для *WW* подтверждают указанное выше свойство этих графов. Рассмотрим целочисленные интервалы $I_1 = [WW(S_n), WW(S'_n)]$ и $I_2 = [WW(S_n'), WW(S_n'')]$. Тогда длины первых двух интервалов нереализуемых значений с левой стороны E_n равны $|I_1| = 3(n-3) - 1 = 3n - 10$ и $|I_2| = 3(n-5) - 1 = 3n - 16$. С правой стороны E_n длины соответствующих интервалов $I_3 = [WW(P''_n), WW(P_n')]$ и $I_4 = [WW(P_n'), WW(P_n)]$ определяются полиномами второй степени от n: $|I_3| = (n+2)(n-5)/2 - 1 = (n^2 - 3n - 12)/2$ и $|I_4| = (n+2)(n-3) - 1 = (n^2 - n - 8)/2$.

Отметим, что для нахождения деревьев с совпадающими значениями гиперииндекса Винера в [11] использовался метод случайного порождения матриц смежности деревьев. В ходе экспериментальных расчетов было найдено 50 деревьев с числом вершин $n = 20$ и заданным значением $WW = 1895$ за 14 ч работы компьютера типа IBM PC AT-486. Из полученных численных данных по распределению числа деревьев следует, что точный размер класса вырождения для $WW = 1895$ равен 425.

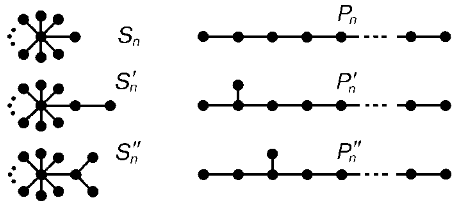

Рис. 2. Деревья с экстремальными значениями гипериндекса Винера

Прикладная ценность полученных результатов для задач структурной химии состоит в четкой формулировке границ применимости гиперииндекса Винера к однозначному описанию химических соединений древовидной структуры и в оценке вырождения индекса для классов деревьев с числом вершин $n \geq 27$. Кроме того, теоретическое нахождение интервалов нереализуемых значений индекса может оказаться полезным при решении так называемой обратной задачи в проблеме связи структура — свойство [19]. В такой задаче требуется по гипотетическому значению индекса, полученному, например, из корреляционного уравнения, построить соответствующие молекулярные графы. Если такое значение попадает в интервал нереализуемых значений, то нужных молекулярных графов не существует. При этом можно указать ближайшее значение, для которого имеет смысл искать молекулярные структуры.

В заключение сделаем одно замечание, касающееся более широкого класса инвариантов. Используя способ перехода от индекса к гипериндексу Винера (см. формулу (2)), можно формально построить семейство инвариантов:

$$W_k(G) = C(k) \sum_{x<y} \sum_{i=1}^{k} d(x,y|G)^i ,$$

где $C(k)$ — константа. Тогда $W_1(G) = W(G)$ и $W_2(G) = WW(G)$ при $C(1) = 1$ и $C(2) = 1/2$. Нетрудно понять, что дискриминирующая способность W_k возрастает (по крайней мере, не убывает) с увеличением k, так как область значений инвариантов расширяется. Для иллюстрации приведем значения этих инвариантов на звезде S_n и цепи P_n (для простоты полагаем $C(k) = 1$):

$$W_3(S_n) = 2(n-1)(7n-11),$$

$$W_4(S_n) = 2(n-1)(15n-26),$$

$$W_3(P_n) = \frac{1}{30} n(n-1)(n+1)(3n^2+5n+8),$$

$$W_4(P_n) = \frac{1}{30} n(n-1)(n+1)(n+2)(2n^2-n+4).$$

Укажем для инварианта W_k нижнюю оценку мощности максимального класса вырождения. В теории графов и математической химии вектор-инвариант $D(G) = (d_0, d_1,..., d_n)$, где d_i — число неупорядоченных пар вершин графа G, находящихся друг от друга на расстоянии i [20 — 22]. Обозначим $N(TI)$ размер максимального класса вырождения для топологического индекса TI на любом произвольно заданном множестве графов. Тогда выполняется простая оценка $N(W_k) \geq N(D)$ для любого k. Таким образом, инвариант D дает нижнюю оценку мощности максимального класса вырождения для W_k вне зависимости от роста k.

СПИСОК ЛИТЕРАТУРЫ

1. *Randić M.* // Chem. Phys. Lett. – 1993. – **211**. – P. 478 – 483.
2. *Randić M., Guo X., Oxley T., Krishnapriyan H.* // J. Chem. Inf. Comput. Sci. – 1993. – **33**. –P. 709 – 716.
3. *Randić M., Guo X., Oxley T. et al* // Ibid. – 1994. – **34**. – P. 361 – 367.
4. *Lukovits I., Linert W.* // Ibid. – P. 899 – 902.
5. *Lukovits I.* // Ibid. – P. 1079 – 1081.
6. *Klein D.J., Lukovits I., Gutman I.* // Ibid. – 1995. – **35**. – P. 50 – 52.
7. *Linert W., Renz F., Kleestorfer K., Lukovits I.* // Computers Chem. – 1995. – **19**. – P. 395 – 401.
8. *Linert W., Kleestorfer K., Renz F., Lukovits I.* // J. Mol. Struct. (Theochem). – 1995. – **337**. – P. 121 – 127.

9. *Zhu H.Y., Klein D.J., Lukovits I.* // J. Chem. Inf. Comput. Sci. – 1996. – **36**. – P. 420 – 428.
10. *Linert W., Lukovits I.* // Commun. Math. Chem. (MATCH). – 1997. – N 35. – P. 65 – 74.
11. *Gutman I., Linert W., Lukovits I., Dobrynin A.A.* // J. Chem. Inf. Comput. Sci. – 1997. – **37**. – P. 349 – 354.
12. *Wiener H.* // J. Amer. Chem. Soc. – 1947. – **69**. – P. 17 – 20.
13. *Gutman I., Polansky O.E.* Mathematical Concepts in Organic Chemistry. – Berlin: Springer-Verlag, 1986.
14. *Применение* теории графов в химии / Под. ред. Н.С. Зефирова, С.И. Кучанова. – Новосибирск: Наука, 1988.
15. *Станкевич М.И., Станкевич И.В., Зефиров Н.С.* // Усп. химии. – 1988. – **57**. – С. 337 – 366.
16. *Razinger M., Chretien J.R., Dubois J.* // J. Chem. Inf. Comput. Sci. – 1985. – **25**. – P. 23 – 27.
17. *Gutman I., Šoltés L.* // Z. Naturforsch. – 1991. – **46a**. – P. 865 – 868.
18. *Gutman I, Luo Y.L., Lee S.L.* // J. Chin. Chem. Soc. – 1993. – **40**. – P. 195 – 198.
19. *Баскин И.И., Гордеева Е.В., Девдариани Р.О. и др.* // Докл. АН СССР – 1989. – **307**. – С. 613 – 617.
20. *Buckley F.* // Lect. Notes Math. – 1984. – **1073**. – P. 179 – 192.
21. *Hosoya H.* // Discrete Appl. Math. – 1988. – **19**. – P. 239 – 257.
22. *Halberstam F.Y., Quintas L.V.* Distance and Path Degree Sequences for Cubic Graphs. –N.Y.: Pace University, Mathematical Department, 1982.

Институт математики им. С.Л. Соболева
СО РАН,
630090 Новосибирск
пр. Коптюга, 4
E-mail: dobr@math.nsc.ru
Научный факультет университета
г. Крагуевац (Югославия)
E-mail: gutman@knez.uis.kg.ac.yu
Новосибирский институт органической
химии им. Н.Н. Ворожцова СО РАН
Научно-технический центр химической
информатики
630090 Новосибирск
пр. Акад. Лаврентьева, 9
E-mail: piootlukh@nioch.nsc.ru

Статья поступила
13 октября 1997 г.