

ИНФОРМАЦИОННАЯ ОЦЕНКА ХИМИЧЕСКИХ СТРУКТУР МЕТОДАМИ АРИФМЕТИЧЕСКИХ ГРАФОВ

Ю. Г. ГРИГОРЬЯН, А. В. ПЕТРОСЯН

Одной из актуальных проблем современной химии – проблема установления связи между структурой и свойствами молекул [1]. Такая связь позволяет получить информационную оценку структуры химического графа молекулы, как меры её неоднородности или разнообразия. Например, из двух графов G_1 и G_2 с равным числом вершин, граф G_1 является более сложным, чем G_2 , если множество вершин графа G_1 разбивается на большее число классов эквивалентности, чем в случае графа G_2 . Следует отметить, что информационное содержание графа неоднозначно и зависит от способа, с помощью которого на множестве было определено отношение эквивалентности.

Особый интерес для химиков представляет изоморфизм и кодирование графов [2] - направление связания с теорией арифметических графов [3]. Для удобства изложения последующих вопросов приведём некоторые определения связанные с понятием арифметического графа. Согласно [3,4,5], любой конечный граф может быть закодирован с помощью целых чисел и представлен в некоторой стандартной теоретико-числовой форме, известной как арифметический граф.

Статистические свойства арифметических графов, установленные в [6] создали соответствующие предпосылки для изучения информационных свойств химических графов.

Пусть даны два конечных множества целых положительных чисел:

$$N = \{n_1, n_2, \dots, n_s\} \text{ и } M = \{m_1, m_2, \dots, m_k\} \quad (1)$$

Определение 1. Арифметическим графом называется пара $G(N, M)$, где N -множество вершин графа, (n_i, n_j) - его ребро, если $n_i + n_j \in M$; M - порождающее множество, а число $n_i + n_j = m_r$ – вес ребра (n_i, n_j) . В работе [4] было показано,

что понятие арифметического графа (а. г.) и графа в классическом смысле [6] эквивалентны. Для отличия от обычных графов, а.г. обозначим $G(N/M)$.

Определение 2. Количество ребер графа $G(N/M)$, порождаемых числом $n_i + n_j = m_r$, назовём частотой элемента m_r и обозначим $v(m_r)$. Другими словами, $v(m_r)$ - число решений уравнения $n_i + n_j = m_r$, $i \neq j$ над множеством N и поэтому $v = f(m_r, N)$. Отсюда число ребер q графа $G(N/M)$ можно выразить формулой:

$$q = \sum_{i=1}^k V(m_i) \quad (2)$$

Из (1),(2) и определения 2 следует, что количество классов эквивалентности над ребрами а.г. $G(N/M)$ и два ребра $(n_i, n_j), (n'_i, n'_j)$ принадлежат подграфу $G_s(N/\{m_s\})$ графа $G(N/M)$, если

$$n_i + n_j = n'_i + n'_j \quad (3)$$

Определение 3. Граф $G(N/M)$ изоморфен графу $G(N'/M')$, если существует взаимно-однозначные соответствие между N и N' такое, что

$$m_r = n_i + n_j \in M \Leftrightarrow \varphi(n'_i) + \varphi(n'_j) \in M' \quad (4)$$

Из определения 1 возникла идея введения специального отношения эквивалентности над ребрами химического графа под которым понимается молекулярная структура, где атомы представляются вершинами, а ковалентные химические связи-ребрами. Такой граф описывает связанность атомов в молекуле независимо от метрических свойств пространства, в котором она существует. Поэтому химические графы являются топологическими представлениями молекулярных структур. В дальнейшем будем рассматривать помеченные химические графы (ПХГ) в котором атомы помечены индексами из натуральных

чисел $\{1,2,\dots,p\}$ произвольным образом. При таком подходе множество ребер (ПХГ) можно разбить на подклассы, состоящие из ребер одного типа (a,b) , где a, b - обозначения атомов. В частности, существуют химические графы, состоящие всего из одного типа ребра. Например, для метана (CH_4) имеются 4 ребра типа (CH) и соответствующий класс эквивалентности - единственный $K(\text{CH})$. Предположим L – помеченный химический граф.

Определение 4. Два ребра $(a_i, b_j), (c_s, d_t)$ принадлежащие L , называются эквивалентными, если пары соответствующих букв $(a,b), (c,d)$ (без индексов) удовлетворяют условию:

$$\{a,b\} = \{c,d\}, \quad (5)$$

где a,b,c,d рассматриваются как некоторые элементы из множества атомов.

Для удобства понимания последующего материала приведём пример химического графа бутанол-1 [Рис.1] на котором будет иллюстрироваться предложенный метод.

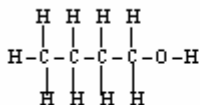


Рис.1

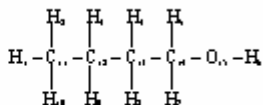


Рис.2

На Рис.2 представлен (ПХГ) бутанол-1 с числом вершин $p=15$ в котором число атомов водорода 10, углерода 4, кислорода 1. Согласно определению 4, число классов эквивалентности в L равно 4 (по числу типов ковалентных связей: (CH) , (OH) , (CO) , (CC)). Соответствующие классы эквивалентности обозначим: $K_{(\text{CH})}$, $K_{(\text{OH})}$, $K_{(\text{CO})}$, $K_{(\text{CC})}$. Общее число ребер q в L равно 14.

$$q = |K_{(\text{CH})}| + |K_{(\text{OH})}| + |K_{(\text{CO})}| + |K_{(\text{CC})}| = 9 + 1 + 1 + 3 = 14$$

В (ПХГ) L зафиксируем некоторую вершину V_i ($i = \overline{1,p}$) и ее

элементы U_t из окрестности первого порядка (рис.3)

$$U_1, U_2, \dots, U_t, \dots, U_\ell \quad (t = \overline{1, \ell}) \quad (6)$$

Каждому ребру (V_i, U_t) пучка (рис.3) припишем число ребер $|K_{(V_i, U_t)}|$, принадлежащих классу $K_{(V_i, U_t)}$ и найдем M_i :

$$M_i = \sum_{t=1}^{\ell} |K(V_i, U_t)|$$

Определение 5. Структурной массой вершин V_i назовем число

$$\overline{M}_i = |V_i| \sum_{t=1}^{\ell} |K(V_i, U_t)| \quad (7)$$

где V_i - атомный вес элемента $V_i \in L$

Определение 6. Структурной массой молекулы (ПХГ) L называется величина

$$M = \sum_{i=1}^p \overline{M}_i \quad (8)$$

где p - число вершин (ПХГ) L .

Пусть V_i некоторая вершина (ПХГ) и ℓ - степень вершины $V_i (\ell = \text{deg} V_i)$. Зафиксируем V_i и рассмотрим пучок-граф на Рис.3. Каждому ребру $(V_i, U_t) \in K(V_i, U_t)$, $t = \overline{1, \ell}$ на Рис.3 припишем число $|K(V_i, U_t)|$ элементов множества $K(V_i, U_t)$. В результате, получим нагруженный пучок состоящий из ℓ -ребёр.

Вероятности распределения структурной массы молекулы (ПХГ) по вершинам V_i выразится формулой:

$$P_i = \frac{\overline{M}_i}{M} \quad (9)$$

На основании (9) определяем P_i и количество информации H по формуле Шеннона:

$$H = -\sum_{i=1}^s P_i \log_2 P_i \quad (10)$$

Для иллюстрации предложенного метода при исследовании количественных корреляций "структура химического графа спиртов растворимость в воде", приведем пример, взятый из [8].

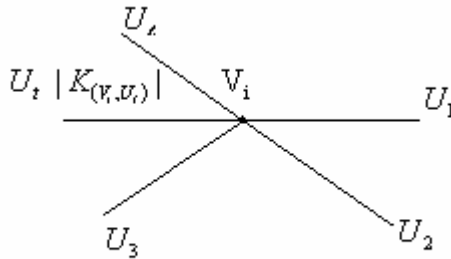


Рис.3

В таблице 1 указаны показатели растворимости в воде в мольных долях 51 спиртов и соответствующие им величины количества информации H в битах (10), полученные расчётным путём. Корреляционный коэффициент $R_{3,4}=0,91$ является достаточно высоким. Число классов эквивалентности равно четырём -

$$K_{(CH)}, K_{(OH)}, K_{(CO)}, K_{(CC)}.$$

Аналогичные расчёты были выполнены по данным для другой группы углеводородных соединений (таблица 2), удовлетворяющих определенным физическим условиям "таяния" и "кипения". Корреляционные коэффициенты $R_{5,7}=0,97$, $R_{6,7}=0,98$ этого примера выше, чем для примера 1, что объясняется единственностью класса эквивалентности - $K_{(CH)}$. Полученные достаточно высокие корреляционные коэффициенты показывают перспективность разбиения структуры химического графа на классы эквивалентности по типам ребер и создают условия для решения других задач системного характера. Следует отметить, что предложенный метод является общим и основан на теории

арифметических графов [3].

Ключевые слова: граф, структура, корреляция, молекула, эквивалентность, информация.

ЛИТЕРАТУРА

1. **Химические** приложения топологии и теории графов /Редактор Д. Кинг. - М.: „Мир“, 1987. - 560с.
2. **Lasky J., Henderson J. W.** //Republic Scence, 1982, Nov.-79 p.
3. **Григорьян Ю.Г.** Теоретико-числовые, алгебраические и геометрические основы арифметических графов: Автореф. дис. д-ра физ.- мат. наук / РАН. Вычисл. Центр. – М.: 1992. – 135с.
4. **Григорьян Ю.Г., Маноян Г.К.** Некоторые вопросы арифметической интерпретации неориентированных графов // Кибернетика. – 1977. – №3. – с. 129-131
5. **Григорьян Ю.Г.** Геометрия арифметических графов // Кибернетика – 1982. – №4. – с. 1- 4.
6. **Григорьян Ю.Г.** Классификация и статистические свойства арифметических графов // Кибернетика – 1979. – №6. – с. 9-12.
7. **Харари Ф.,** Теория графов. М.: „Мир“, -1973.
8. **Магнусон В., Харрис Д., Бейсак С.** Химические приложения топологии и теории графов, М.:„Мир“,1987, с.206-221.
9. **Григорьян Ю.Г.** Группа арифметических автоморфизмов простых циклов // Кибернетика. – 1990. – №4, с.9-15.

Յու.Գ.ԳՐԻԳՈՐՅԱՆ, Ա.Վ.ՊԵՏՐՈՍՅԱՆ

ՔԻՄԻԱԿԱՆ ԿԱՌՈՒՑՎԱԾՔՆԵՐԻ ՎԵՐԱԲԵՐՅԱԼ ՏԵՂԵԿԱՏՎՈՒԹՅԱՆ ԳՆԱՀԱՏՈՒՄԸ ԹՎԱԲԱՆԱԿԱՆ ԳՐԱՖՆԵՐԻ ՄԵԹՈԴՈՎ

Ամփոփում

Հոդվածում ներկայացված է թվաքանակային (կողավորված) գրաֆների տեսության վրա հիմնված քիմիական կառուցվածքների տեղեկատվության գնահատման յուրահատուկ մեթոդ: Մոլեկուլյար քիմիական կառուցվածքների կովալենտ կապերը տրվում են թվաքանակային գրաֆների տեսքով: Այդպիսի մեթոդով կառուցվում և Շենոնի դասական բանաձևով հաշվարկվում է մի շարք նյութերի տեղեկատվության քանակը: Նախագծվել է ավտոմատացված համակարգ, որը հաշվարկում է տրված նյութի տեղեկատվության ծավալը: Արդյունքում ստացվել են մի շարք քիմիական նյութերի տեղեկատվական քանակները, որոնք ներկայացվել են աղյուսակների ձևով: Հաշվարկվել են նաև կորելյացիաները նյութերի ֆիզիկա-քիմիական և տեղեկույթի քանակների միջև: Ստացված գործակիցները բավականին բարձր են և տատանվում են 0,91- 0,98 մեծությունների սահմաններում: Ներկայացված մեթոդն ընդհանուր բնույթ է կրում և կարող է օգտագործվել այնպիսի խնդիրների լուծման ընթացքում, ինչպիսիք են նյութերի մոլեկուլների ֆիզիկա-քիմիական և կենսաբանական հատկությունների բացահայտումը, նկարագրությունն ու գնահատումը:

Առանցքային բառեր. *գրաֆ, կառուցվածք, մոլեկուլ, տեղեկատվություն, կորելյացիա:*

Yu. G. GRIGORYAN, A.V. PETROSYAN

**EVALUATION OF THE INFORMATION ABOUT
CHEMICAL STRUCTURES BY ARITHMETIC GRAPHS**

Summary

The article proposes a method for assessing the information structures of chemical compounds on the basis of the theory of arithmetic (coded) graphs. In the chemical graph of the molecule a special relationship of equivalence between the covalent bonds similar to ribs arithmetic graphs is established. On this basis, the corresponding equivalence classes are based on the chemical graph of the molecule and calculated the amount of information on its structure to Shannon. By this method, a program is made for specific calculations of certain chemicals. We calculate the correlation between their chemical, physical properties and the amount of information in the structure of the molecule. The resulting correlation coefficients are ranged 0.91-0.98 and are sufficiently high. The proposed method is general and can be applied to life a variety of problems for the quantitative description of the molecules in the study of their physical, chemical and biological properties.

Key words: *graph, structure, correlation, information, molecule.*

ПРИЛОЖЕНИЕ**Таблица 1**

N	Название	Раствори- мость (-logX)	Кол-во информации H /бит /
1	2	3	4
1	Бутанол-1	1.750	2.518
2	2-Метилпропанол	1.743	2.498
3	Бутанол-2	1.724	2.490
4	Пентанол-1	2.332	2.800
5	3-Метилбутанол	2.254	2.782
6	2-Метилбутанол	2.207	2.783
7	Пентанол-2	2.025	2.775
8	Пентанол-3	1.961	2.775
9	3-Метилбутанол-2	1.926	2.761
10	2-Метилбутанол-2	1.608	2.732
11	2,2-Диметилпропанол-1	2.030	2.997
12	Гексанол-1	2.957	3.033
13	Гексанол-2	2.612	3.015
14	Гексанол-3	2.542	3.015
15	3-Метилпентанол-3	2.109	2.981
16	2-Метилпентанол-2	2.333	2.981
17	2- Метилпентанол -3	2.445	3.004
18	3- Метилпентанол -2	2.458	3.004
19	2,3-Диметилбутанол-2	2.118	2.970
20	3,3- Диметилбутанол	2.870	2.999
21	3,3- Диметилбутанол-2	2.359	2.979
22	4- Метилпентанол	2.737	3.022
23	4- Метилпентанол -2	2.534	3.004
24	3-Этилбутанол	2.956	3.022
25	Циклогексанол	2.164	2.968
26	Гептанол-1	3.554	3.033
27	2-Метилгексанол-2	2.820	3.194
28	3-Метилгексанол-3	2.729	3.213
29	3-Этилпентанол-3	2.578	3.194

*Продолжение
Таблица 1*

N	Название	Раствори- мость (-logX)	Кол-во информации H /бит /
1	2	3	4
30	2,3-Диметилпентанол-2	2.615	2.970
31	2,3- Диметилпентанол -3	2.588	3.163
32	2,4- Диметилпентанол -2	2.678	3.186
33	2,4- Диметилпентанол -3	2.962	3.205
34	2,2- Диметилпентанол -3	2.893	3.180
35	Гептанол-3	3.132	3.222
36	Гептанол-4	3.133	3.222
37	Октанол-1	4.091	3.601
38	2,2,3-Триметилпентанол-3	3.018	3.356
39	Октанол-2	3.811	3.404
40	2-Этилгексанол	3.915	3.409
41	Нонанол-1	4.745	3.577
42	Нонанол-2	4.490	3.566
43	Нонанол-3	4.402	3.566
44	Нонанол-4	4.330	3.566
45	Нонанол-5	4.240	3.566
46	2,6-Диметилпентанол-3	4.253	3.213
47	3,5-Диметилпентанол-4	4.046	3.213
48	1,1-Диметилпентанол	4.165	3.545
49	7-Метиоктанол	4.240	3.570
50	3,5,5-Триметилгексанол	4.251	3.550
51	Деканол-1	5.444	3.721

Таблица 2

N	Химическая структура	Название	Состояние	Температура плавления °С	Температура кипения °С	Кол-во информации Н / бит /
1	2	3	4	5	6	7
1	CH ₄	Метан	Газ	-182	-164	0,5451
2	C ₂ H ₆	Этан	Газ	-183	-89	1,4936
3	C ₃ H ₈	Пропан	Газ	-190	-42	1,5895
4	C ₄ H ₁₀	Бутан	Газ	-138	-1	2,8103
5	C ₅ H ₁₂	Пентан	Жидкость	-130	36	2,7367
6	C ₆ H ₁₄	Гексан	жидкость	-95	69	2,9903
7	C ₇ H ₁₆	Гептан	жидкость	-91	98	3,2057
8	C ₈ H ₁₈	Октан	жидкость	-57	126	3,3930
9	C ₉ H ₂₀	Нонан	жидкость	-51	151	3,5586
10	C ₁₀ H ₂₂	Декан	жидкость	-30	174	3,7071
11	C ₁₁ H ₂₄	Ундекан	жидкость	-26	196	3,8416
12	C ₁₂ H ₂₆	Додекан	жидкость	-10	216	3,9748
13	C ₁₃ H ₂₈	Тридекан	жидкость	-6	235	4,0780
14	C ₁₄ H ₃₀	Тетрадекан	жидкость	6	254	4,1831
15	C ₁₅ H ₃₂	Пентадекан	жидкость	10	271	4,2810
16	C ₂₀ H ₄₂	Эйкозан	твердый	37	343	4,7086
17	C ₃₀ H ₆₂	Триадоктан	твердый	66	450	5,2694
18	C ₄₀ H ₈₂	Тетрадонтан	твердый	80	450	5,6814
19	C ₅₀ H ₁₀₂	Пентадонтан	твердый	93	450	6,0015