

ХЕМОИНФОРМАТИКА: ПРИЛОЖЕНИЯ ИНФОРМАТИКИ В АНАЛИЗЕ ХИМИЧЕСКИХ СТРУКТУР (на примере сульфида кадмия)*

Максимов А. Г.^{1,2}, младший научный сотрудник, agm@dscs.pro
Завалишин А. Д.^{1,2}, младший научный сотрудник, adz@dscs.pro
Абрамов М. В.^{1,2}, кандидат технических наук, ✉ mva@dscs.pro
Тулупьев А. Л.^{2,1}, доктор физико-математических наук, alt@dscs.pro

¹ Санкт-Петербургский институт информатики и автоматизации Российской академии наук,
14 линия, 39, 199178, Санкт-Петербург, Россия

² Санкт-Петербургский государственный университет,
Университетская наб., 7–9, 199034, Санкт-Петербург, Россия

Аннотация

Статья направлена на демонстрацию практического приложения теории графов как подраздела теоретических основ информатики в решении одной из междисциплинарных задач — описании структуры молекулы сульфида кадмия с применением методов и индексов хемиоинформатики. В статье представлены результаты вычислений atom-bond connectivity индекса (ABC), геометрического и арифметического индекса GA, обобщенного индекса Рандича, GA5 и ABC4, загребских индексов для химического графа сульфида кадмия (CdS). Топологические индексы для сульфида кадмия рассматриваются впервые, хотя сама по себе задача расчета данных индексов не нова. Актуальность результатов подчеркивается тем, что сульфид кадмия широко используют в различных областях, таких как оптоэлектроника, фотоприемники, фоторезисторы и т. д.

Ключевые слова: топологический индекс, сульфид кадмия, хемиоинформатика, теория графов, молекулярный дескриптор, индекс рандича, загребский индекс.

Цитирование: Максимов А. Г., Завалишин А. Д., Абрамов М. В., Тулупьев А. Л. Хемиоинформатика: приложения информатики в анализе химических структур (на примере сульфида кадмия) // Компьютерные инструменты в образовании. 2019. № 4. С. 44–54. doi:10.32603/2071-2340-2019-4-44-54

1. ВВЕДЕНИЕ

Часто математика и информатика используются в приложении к исследованиям различных областей знаний [20–24]. Один из примеров таких приложений — хемиоинформатика — относительно недавно появившаяся область исследований. Среди прочих про-

*Работа выполнена в рамках проекта по государственному заданию СПИИРАН № 0073-2019-0003, при финансовой поддержке РФФИ — проект № 18-01-00626.

блема анализа структуры молекулы конкретного вещества может относиться к задачам хемоинформатики. Представляется полезным рассмотреть на примере ещё нерешенных задач по анализу структуры молекулы конкретного вещества подходы, в этой области используемые, чтобы привлечь внимание студентов, аспирантов и исследователей в области информатики к целому пласту неисследованных проблем.

Итак, существует много соединений органической и неорганической химии, изучение свойств и структуры которых очень важно для промышленной, фармацевтической и лабораторной химии [1]. Изучение этих свойств и структуры видится перспективным для поиска и оценки вариантов использования вещества [2]. Важно, что между структурой молекулы некоторого вещества и его химическими свойствами существует прямая связь [3]. То есть изучение структуры молекулы вещества даёт информацию о его химических свойствах. Обозначенные молекулярные структуры можно представить в виде графов, у которых в вершинах будут находиться атомы, а ребра будут отображать связи между ними. Это позволит использовать для изучения структуры молекулы вещества методы теории графов.

Ключевая характеристика структуры молекулы вещества — её топологический индекс. Топологический индекс — это инвариант молекулярного графа, числовая величина, связанная с химическим строением, которая указывает на корреляцию химической структуры со многими физическими, химическими свойствами и биологической активностью [1]. Иными словами, зная топологический индекс вещества, мы имеем представление о его химических свойствах. Таким образом, актуальной видится задача расчёта топологических индексов различных химических веществ.

Научная цель данной статьи — описать структуру молекулы сульфида кадмия, применяя методы и индексы хемоинформатики. Дидактическая цель материала заключается в том, чтобы на конкретном примере продемонстрировать практическое приложение теории графов как подраздела теоретических основ информатики в решении одной из междисциплинарных задач.

В статье рассмотрен пример расчета некоторых топологических индексов для молекулярного графа сульфида кадмия (CdS) с гексагональной решёткой [1, 6]. Топологические индексы для сульфида кадмия рассматриваются впервые, хотя сама по себе задача расчета данных индексов не нова, так, например, в работах [1, 4] были рассмотрены индексы оксидов меди. Сульфид кадмия широко используют в оптоэлектронике как в фотоприёмниках, так и в солнечных батареях [7, 8]. Это вещество применяют для изготовления фоторезисторов (приборов, электрическое сопротивление которых меняется в зависимости от освещенности) [8]. Красители на основе данного соединения обладают высокой термо- и светостойкостью [7, 8]. Также сульфид кадмия применяется для регистрации элементарных частиц и гамма-излучения и в качестве рабочего тела твердотельных лазеров [7, 8]. На основе сульфида кадмия создают наноструктурированные материалы (квантовые точки, нанопроволоки, нанотрубки и т. д.), то есть обычные соединения, синтезированные в искусственно созданных границах «ограниченной геометрии» [6–8]. Такие материалы находят применение в медицине и биологии в качестве люминесцентных меток [8].

2. ПОСТАНОВКА ЗАДАЧИ

Одним из наиболее ранних и важных топологических индексов, базирующихся на степенях вершин, был индекс Рандича [11]:

$$R_{-\frac{1}{2}}(G) = \sum_E \frac{1}{\sqrt{d_u d_v}}.$$

В 1988 году был предложен двумя исследователями независимо обобщенный индекс Рандича [12, 13]:

$$R_\alpha(G) = \sum_E (d_u d_v)^\alpha.$$

Еще одним важным топологическим индексом является первый Загребский индекс [14]:

$$M_1(G) = \sum_E (d_u + d_v).$$

Несколько позже был введен также и второй Загребский индекс [15]:

$$M_2(G) = \sum_E d_u d_v.$$

Существенное значение также имеет индекс атомной связности [16]:

$$ABC(G) = \sum_E \sqrt{\frac{d_u + d_v - 2}{d_u d_v}}.$$

Известна в том числе его версия, называемая ABC_4 [17]:

$$ABC_4(G) = \sum_E \sqrt{\frac{S_u + S_v - 2}{S_u S_v}}.$$

Наконец, геометрический арифметический индекс определяется следующим образом [18]:

$$GA(G) = \sum_E \frac{2\sqrt{d_u d_v}}{d_u + d_v},$$

где

$$S_u = \sum_{uv \in E} d_v, S_v = \sum_{uv \in E} d_u,$$

а также его версия GA_5 [19]:

$$GA_5(G) = \sum_E \frac{2\sqrt{S_u S_v}}{S_u + S_v}.$$

3. РАСЧЕТЫ ТОПОЛОГИЧЕСКИХ ИНДЕКСОВ

Молекулярный граф сульфида кадмия представляет собой пространственную шестиугольную решетку [9, 10], как на рис. 1.

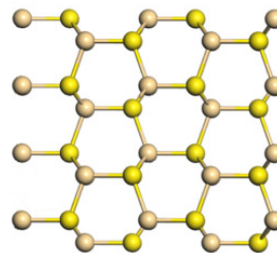
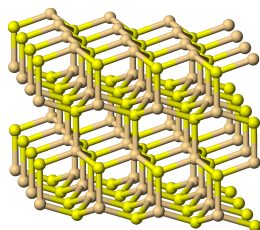


Рис. 1. Структура SdS (желтым цветом выделены атомы S)

Для удобства дальнейших выкладок обозначим за m «высоту» решетки, то есть максимальное число шестиугольников в столбце, и обозначим за n число столбцов высоты m (столбцы высоты m и $m - 1$ чередуются, начиная и заканчивая высотой m). Сразу оговоримся, что общее число столбцов тогда $2n - 1$. Посчитаем мощности вершин и ребер. $|V| = (4m + 2)n = 4mn + 2n$, $|E| = 4mn + (m + 1)n + m(n - 1) = 6mn + n - m$. Посчитаем также число вершин с конкретной степенью. Обозначим множество вершин степени n как D_n , под d_n будем понимать представителя этого множества. Тогда $|D_2| = 2n \cdot 2 + 2m = 2m + 4n$, $|D_3| = |V| - |D_2| = 4mn - 2n - 2m$. В таблице 1 отражена информация о числе ребер с фиксированным числом соседей у каждого из концов.

Таблица 1. Распределение ребер графа CdS по степеням концевых вершин

d_u, d_v	Количество
2,2	$4 + 2n$
2,3	$4n + 4m - 8$
3,3	$6mn - 5n - 5m + 4$

Произведем проверку: $\sum = 6mn - 5n - 5m + 4 + 4n + 4m - 8 + 4 + 2n = 6mn + n - m = |E|$.

После выполненной предварительной подготовительной работы можно сформулировать ряд теорем.

Теорема 1. ABC индекс при $G=CdS$ и $m, n > 1$ определяется следующим выражением: $ABC(G) = 4mn + m(2\sqrt{2} - \frac{10}{3}) + n(3\sqrt{2} - \frac{10}{3}) + (\frac{8}{3} - 2\sqrt{2})$.

Доказательство. Как было представлено выше, ABC индекс определяется следующей формулой [16]:

$$ABC(G) = \sum_E \sqrt{\frac{d_u + d_v - 2}{d_u d_v}}$$

Из определения ABC индекса и данных из таблицы 1 имеем:

$$ABC(G) = (4 + 2n) \frac{\sqrt{2}}{2} + (4n + 4m - 8) \frac{\sqrt{2}}{2} + (6mn - 5n - 5m + 4) \frac{2}{3}$$

В результате упрощения и группировки имеем:

$$ABC(G) = 4mn + m(2\sqrt{2} - \frac{10}{3}) + n(3\sqrt{2} - \frac{10}{3}) + (\frac{8}{3} - 2\sqrt{2}). \blacksquare$$

Следующая теорема показывает значение геометрического арифметического индекса молекулярного графа CdS.

Теорема 2. Геометрический арифметический индекс для графа CdS равен $GA(G) = 5mn + m(\frac{8\sqrt{6}}{5} - 5) + n(\frac{8\sqrt{6}}{5} - 3) + (8 - \frac{16\sqrt{6}}{5})$.

Доказательство. Индекс GA по определению равен [18]

$$GA(G) = \sum_E \frac{2\sqrt{d_u d_v}}{d_u + d_v}$$

С помощью данных из таблицы 1 получаем

$$GA(G) = (4 + 2n) + (4n + 4m - 8) \frac{2\sqrt{6}}{5} + (6mn - 5n - 5m + 4)$$

В результате упрощения и группировки имеем:

$$GA(G) = 5mn + m\left(\frac{8\sqrt{6}}{5} - 5\right) + n\left(\frac{8\sqrt{6}}{5} - 3\right) + \left(8 - \frac{16\sqrt{6}}{5}\right). \blacksquare$$

Следующая теорема показывает значение обобщенного индекса Рандича молекулярного графа CdS.

Теорема 3. Обобщенный индекс Рандича для графа CdS равен

$$R_\alpha(G) = \begin{cases} 54mn + 21m + 13n + 4 & \alpha = 1 \\ \frac{2}{3}mn + \frac{1}{9}m + \frac{11}{18}n + \frac{1}{9} & \alpha = -1 \\ 18mn + m(4\sqrt{6} - 15) + n(4\sqrt{6} - 11) + (20 - 8\sqrt{6}) & \alpha = \frac{1}{2} \\ 2mn + m\left(\frac{2\sqrt{6}}{3} - \frac{5}{3}\right) + n\left(\frac{2\sqrt{6}}{3} - \frac{2}{3}\right) + \left(\frac{10}{3} - \frac{4\sqrt{6}}{3}\right) & \alpha = -\frac{1}{2} \end{cases}$$

Доказательство. Индекс R_α по определению равен [12]

$$R_\alpha(G) = \sum_E (d_u d_v)^\alpha.$$

При $\alpha = 1$ с помощью данных из таблицы 1 получаем

$$R_1(G) = (4 + 2n)4 + (4n + 4m - 8)6 + (6mn - 5n - 5m + 4)9.$$

В результате упрощения и группировки имеем:

$$R_1(G) = 54mn + 21m + 13n + 4.$$

При $\alpha = -1$ с помощью данных из таблицы 1 получаем

$$R_{-1}(G) = (4 + 2n)\frac{1}{4} + (4n + 4m - 8)\frac{1}{6} + (6mn - 5n - 5m + 4)\frac{1}{9}.$$

В результате упрощения и группировки имеем:

$$R_{-1}(G) = \frac{2}{3}mn + \frac{1}{9}m + \frac{11}{18}n + \frac{1}{9}.$$

При $\alpha = \frac{1}{2}$ с помощью данных из таблицы 1 получаем

$$R_{\frac{1}{2}}(G) = (4 + 2n)2 + (4n + 4m - 8)\sqrt{6} + (6mn - 5n - 5m + 4)3.$$

В результате упрощения и группировки имеем:

$$R_{\frac{1}{2}}(G) = 18mn + m(4\sqrt{6} - 15) + n(4\sqrt{6} - 11) + (20 - 8\sqrt{6}).$$

При $\alpha = -\frac{1}{2}$ с помощью данных из таблицы 1 получаем

$$R_{-\frac{1}{2}}(G) = (4 + 2n)\frac{1}{2} + (4n + 4m - 8)\frac{\sqrt{6}}{6} + (6mn - 5n - 5m + 4)\frac{1}{3}.$$

В результате упрощения и группировки имеем:

$$R_{-\frac{1}{2}}(G) = 2mn + m\left(\frac{2\sqrt{6}}{3} - \frac{5}{3}\right) + n\left(\frac{2\sqrt{6}}{3} - \frac{2}{3}\right) + \left(\frac{10}{3} - \frac{4\sqrt{6}}{3}\right). \blacksquare$$

Следующая теорема показывает значения первого и второго Загребского индекса молекулярного графа CdS.

Теорема 4. Первый и второй Загребские индексы для графа CdS равны

$$M_1(G) = 36mn - 10m - 2n,$$

$$M_2(G) = 54mn + 21m + 13n + 4.$$

Доказательство. Индекс M_1 по определению равен [14]

$$M_1(G) = \sum_E (d_u + d_v).$$

С помощью данных из таблицы 1 получаем

$$M_1(G) = (4 + 2n)4 + (4n + 4m - 8)5 + (6mn - 5n - 5m + 4)6.$$

В результате упрощения и группировки имеем:

$$M_1(G) = 36mn - 10m - 2n.$$

Индекс M_2 по определению равен

$$M_2(G) = \sum_E d_u d_v.$$

Заметим, что он равен обобщенному индексу Рандича R_α с константой $\alpha = 1$, вычисление которого уже было произведено ранее. Таким образом, второй Загребский индекс равен

$$M_2(G) = R_1(G) = 54mn + 21m + 13n + 4. \blacksquare$$

Для дальнейшей работы понадобится следующее определение:

Определение: $S_u = \sum_{uv \in E} d_v$.

Таблицы 2, 3 содержат информацию о числе ребер, концы u, v которых имеют фиксированную величину S_u, S_v .

Таблица 2. Формулы для вычисления

S_u, S_v	Оценка
4,4	0
4,5	$4 \cdot 2$
5,5	$2(n - 2)$
5,7	4
5,8	$4 + 2(n - 2)2$
6,7	$2 \cdot 2(m - 2)$
7,9	$2(m - 1)$
8,8	$(n - 1)2$
8,9	$(n - 1)4$
9,9	$(5(m - 1) + 1 - 6)(n - 1) + (n - 2)(m + 1 - 2)$

Таблица 3. Распределение ребер графа CdS по S_u, S_v конечных вершин

S_u, S_v	Количество
4,4	0
4,5	8
5,5	$2n - 4$
5,7	4
5,8	$4n - 4$
6,7	$4m - 8$
7,9	$2m - 2$
8,8	$2n - 2$
8,9	$4n - 4$
9,9	$6mn - 11n - 7m + 12$

На всякий случай сделаем проверку:

$$\sum = 8 + 2n - 4 + 4 + 4n - 4 + 4m - 8 + 2m - 2 + 2n - 2 + 4n - 4 + 6mn - 11n - 7m + 12 = 6mn + n - m = |E|.$$

Следующая теорема показывает значение четвертого индекса атомной связности молекулярного графа CdS.

Теорема 5. Четвертый индекс атомной связности для графа CdS равен

$$ABC_4(G) = \frac{8}{3}mn + \left(\frac{4\sqrt{2}}{5} + \frac{\sqrt{110}}{20} + \frac{\sqrt{14}}{4} + \frac{\sqrt{30}}{3} - \frac{44}{9}\right)n + \left(\frac{2\sqrt{462}}{21} + \frac{2\sqrt{2}}{3} - \frac{29}{8}\right)m + \left(-\frac{8\sqrt{2}}{5} + \frac{4\sqrt{14}}{7} - \frac{\sqrt{110}}{5} - \frac{3\sqrt{462}}{21} - \frac{2\sqrt{2}}{3} - \frac{\sqrt{14}}{4} - \frac{\sqrt{30}}{3} + \frac{8}{9}\right).$$

Доказательство. Из определения ABC_4 и из таблицы 3 получаем

$$ABC_4(G) = \sqrt{\frac{7}{20}} \cdot 8 + \sqrt{\frac{8}{25}} \cdot (2n - 4) + \sqrt{\frac{2}{7}} \cdot 4 + \sqrt{\frac{11}{40}} \cdot (4n - 4) + \sqrt{\frac{11}{42}} \cdot (4m - 8) + \sqrt{\frac{2}{9}} \cdot (2m - 2) + \sqrt{\frac{7}{32}} \cdot (2n - 2) + \sqrt{\frac{5}{24}} \cdot (4n - 4) + \frac{4}{9} \cdot (6mn - 11n - 7m + 2).$$

В результате упрощения и группировки имеем:

$$ABC_4(G) = \frac{8}{3}mn + \left(\frac{4\sqrt{2}}{5} + \frac{\sqrt{110}}{20} + \frac{\sqrt{14}}{4} + \frac{\sqrt{30}}{3} - \frac{44}{9}\right)n + \left(\frac{2\sqrt{462}}{21} + \frac{2\sqrt{2}}{3} - \frac{29}{8}\right)m + \left(-\frac{8\sqrt{2}}{5} + \frac{4\sqrt{14}}{7} - \frac{\sqrt{110}}{5} - \frac{3\sqrt{462}}{21} - \frac{2\sqrt{2}}{3} - \frac{\sqrt{14}}{4} - \frac{\sqrt{30}}{3} + \frac{8}{9}\right). \blacksquare$$

Следующая теорема показывает значение пятого геометрического арифметического индекса молекулярного графа CdS.

Теорема 6. Пятый геометрический арифметический индекс для графа CdS равен

$$GA_5(G) = 6mn + \left(\frac{16\sqrt{10}}{13} + \frac{48\sqrt{3}}{17} - 7\right)n + \left(\frac{8\sqrt{42}}{13} + \frac{3\sqrt{7}}{4} - 7\right)m + \left(\frac{32\sqrt{5}}{9} + \frac{2\sqrt{36}}{3} - \frac{16\sqrt{10}}{13} - \frac{16\sqrt{42}}{13} - \frac{3\sqrt{7}}{8} - \frac{48\sqrt{3}}{17}\right).$$

Доказательство. Из определения GA_5 и из таблицы 3 получаем

$$GA_5(G) = \sqrt{\frac{80}{81}} \cdot 8 + (2n - 4) + \sqrt{\frac{35}{36}} \cdot 4 + \sqrt{\frac{160}{169}} \cdot (4n - 4) + \sqrt{\frac{168}{169}} \cdot (4m - 8) + \sqrt{\frac{63}{64}} \cdot (2m - 2) + (2n - 2) + \sqrt{\frac{432}{289}} \cdot (4n - 4) + (6nm - 11n - 7m + 2).$$

В результате упрощения и группировки имеем:

$$GA_5(G) = 6mn + \left(\frac{16\sqrt{10}}{13} + \frac{48\sqrt{3}}{17} - 7\right)n + \left(\frac{8\sqrt{42}}{13} + \frac{3\sqrt{7}}{4} - 7\right)m + \left(\frac{32\sqrt{5}}{9} + \frac{2\sqrt{36}}{3} - \frac{16\sqrt{10}}{13} - \frac{16\sqrt{42}}{13} - \frac{3\sqrt{7}}{8} - \frac{48\sqrt{3}}{17}\right). \blacksquare$$

4. ЗАКЛЮЧЕНИЕ

Таким образом, в статье продемонстрировано практическое приложение теории графов как подраздела теоретических основ информатики в решении одной из междисциплинарных задач — описании структуры молекулы сульфида кадмия с применением методов и индексов хемоинформатики. В статье представлены результаты вычислений atom-bond connectivity индекса (ABC), геометрического и арифметического индекса GA, обобщенного индекса Рандича, GA_5 и ABC_4 , загребских индексов для химического графа сульфида кадмия (CdS). Сульфид кадмия широко используют в различных областях, связанных с оптоэлектроникой [7, 8], изготовлением фоторезисторов [8], производством красителей с высокой термо- и светостойкостью [7, 8], применением в наноструктурированных материалах (квантовые точки, нанопроволоки, нанотрубки и т. д.), медицине и биологии в качестве люминесцентных меток [8]. Иными словами, сульфид кадмия (CdS) активно используется в физических, биологических и химических исследованиях, в том числе и производственных исследованиях, поэтому его параметры видятся перспективными для включения в датасет для изучения свойств других соединений в самых разных областях. В дальнейшем описание структур молекул с использованием методов и индексов хемоинформатики может помочь в нахождении новых химических, физических и биологических свойств соединений, которые считались уже изученными, а также в выборе направления развития исследований свойств других соединений.

Список литературы

1. Gao W., Baig A. Q., Khalid W., Farahani M. R. Molecular description of copper (II) oxide // Macedonian Journal of Chemistry and Chemical Engineering. 2017. P. 1–7.
2. Mohajeri A., Alipour M., Ahmadi M. B. A graph theory study on $(ZnS)_n$ ($n = 3 - 10$) nanoclusters // Chemical Physics Letters. 2011. P. 162–166.
3. Taherpour A. A. Quantitative structural relationship and theoretical study of electrochemical properties of $C_{60}@[SWCN(5,5) - Armchair - C_nH_{20}]$ complexes // Chemical Physics Letters. 2009. P. 135–
4. Gao W., Imran M., Siddiqui M. K., Naeem M., Jamil F. Molecular description of copper (I) oxide and copper (II) oxide // Quimica Nova. 2018. P. 874–879.
5. Deng F., Zhang X., Alaeiyan M., Mehboob A., Farahani M. R. Topological Indices of the Pent-Heptagonal Nanosheets VC_5C_7 and HC_5C_7 // Advances in Materials Science and Engineering. 2019. P. 1–12.
6. IARC Monograph: «Cadmium and Cadmium Compounds» Last access Nov. 2005
7. Чопра К., Дас С. Тонкопленочные солнечные элементы: Пер. с англ. с сокращениями // М.: Мир, 1986. 435 с.

8. Тонкие плёнки сульфида кадмия для фотольватки. *Computational nanotechnology*. 2014, 68–73 с.
9. Paul K. *Handbook of Infrared Optical Materials* // CRC Press. 1991.
10. Mohammed B. K. Insight into the origin of magnetism in Iron-doped cadmium sulfide thin films from first principles calculations // *Solid State Communications*. 2017. Vol. 253. P. 10–13.
11. Randić M. On characterization of molecular branching // *J. Amer. Chem. Soc.* 1975. Vol. 97 (23). P. 6609–6615.
12. Amic D., Beslo D., Lucić B., Nikolić S., Trinajstić N. The vertex-connectivity index revisited // *J. Chem. Inf. Comput. Sci.* 1998. Vol. 38. P. 819–822.
13. Bollobas B., Erdos P. Graphs of extremal weights // *Ars Combinatoria*. 1998. № 50. P. 225–233.
14. Gutman I., Das K. C. The first Zagreb index 30 years after. *MATCH Commun. Math. Comput. Chem*, 2004. № 50. P. 83–92.
15. Gutman I., Trinajstić N. Graph theory and molecular orbital's., Total ϕ -electron energy of alternant hydrocarbons. *Chemical Physics Letters*. 1972. № 17. P. 535–538.
16. Estrada E., Torres L., Rodriguez L., Gutman I. An atom-bond connectivity index: Modeling the enthalpy of formation of alkanes // *Indian J. Chem.* 1998. № 37A. P. 849–855.
17. Ghorbani A., Hosseinzadeh M. A. Computing ABC4 index of nanostar Dendrimers // *Optoelectronics and Advanced Materials–Rapid Communications*, 2010. № 4. P. 1419–1422.
18. Vukičević D., Furtula B. Topological index based on the ratios of geometrical and arithmetical means of end-vertex degrees of edges // *J. Math. Chem.* 2009. № 46. P. 1369–1376.
19. Graovac A., Ghorbani M., Hosseinzadeh M. A. Computing fifth geometric-arithmetic index for nanostar Dendrimers // *J. Math. Nanosci.* 2011. № 1. P. 33–42.
20. Абрамов М. В., Тулупьева Т. В., Тулупьев А. Л. Социоинженерные атаки: социальные сети и оценки защищенности пользователей. СПб.: ГУАП, 2018. 266 с.
21. Азаров А. А., Тулупьева Т. В., Суворова А. В., Тулупьев А. Л., Абрамов М. В., Юсупов Р. М. Социоинженерные атаки. Проблемы анализа. СПб.: Наука, 2016. 352 с.
22. Хлобыстова А. О., Абрамов М. В., Тулупьев А. Л., Золотин А. А. Поиск кратчайшей траектории социоинженерной атаки между парой пользователей в графе с вероятностями переходов // *Информационно-управляющие системы*. 2018. № 6. С. 74–81. doi:10.31799/1684-8853-2018-6-74-81
23. Суворова А. В., Смирнова К. Р., Будин Е. А., Тулупьева Т. В., Тулупьев А. Л., Абрамов М. В. Исследовательский проект как инструмент обучения методам анализа текста: предсказание класса поста в социальной сети // *Компьютерные инструменты в образовании*. 2018. № 3. С. 49–64.
24. Kharitonov N. A., Maximov A. G., Tulupuyev A. L. Algebraic Bayesian Networks: Naïve Frequentist Approach to Local Machine Learning Based on Imperfect Information from Social Media and Expert Estimates // *Russian Conference on Artificial Intelligence*. Springer, Cham, 2019. P. 234–244.

Поступила в редакцию 05.11.2019, окончательный вариант — 20.12.2019.

Максимов Анатолий Григорьевич, младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН; студент кафедры информатики СПбГУ, agm@dscs.pro

Завалишин Арсений Дмитриевич, младший научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН; студент кафедры информатики СПбГУ, adz@dscs.pro

Абрамов Максим Викторович, кандидат технических наук, заведующий лабораторией теоретических и междисциплинарных проблем информатики СПИИРАН; доцент кафедры информатики СПбГУ, ✉ mva@dscs.pro

Тулупьев Александр Львович, доктор физико-математических наук, профессор кафедры информатики, СПбГУ; главный научный сотрудник лаборатории теоретических и междисциплинарных проблем информатики СПИИРАН, alt@dscs.pro

Computer tools in education, 2019

№ 4: 44–54

<http://cte.eltech.ru>

[doi:10.32603/2071-2340-2019-4-44-54](https://doi.org/10.32603/2071-2340-2019-4-44-54)

Chemoinformatics: Applications of Computer Science in the Analysis of Chemical Structures (on the Example of Cadmium Sulfide)

Maksimov A. G.^{1,2}, junior researcher, agm@dscs.pro
Zavalishin A. D.^{1,2}, junior researcher, adz@dscs.pro
Abramov M. V.^{1,2}, PhD, senior researcher, mva@dscs.pro
Tulupyev A. L.^{2,1}, PhD, Dc. Sci., professor, alt@dscs.pro

¹Saint Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 39, 14 Line, 199178, Saint Petersburg, Russia

²Saint Petersburg State University, Universitetskaya nab., 7-9, 199034, Saint Petersburg, Russia

Abstract

The article is aimed at demonstrating the practical application of graph theory as a subsection of the theoretical foundations of computer science in solving one of the interdisciplinary problems — describing the structure of the cadmium sulfide molecule using methods and indices of chemoinformatics. The article presents the results of calculations of the atom-bond connectivity index (ABC), of the geometric and arithmetic index GA, of the generalized Randić index, GA5 and ABC4, of the Zagreb indices for the chemical graph of cadmium sulfide (CdS). Topological indices for cadmium sulfide are considered for the first time, although the task of calculating these indices is not new in itself. The relevance of the results is emphasized by the fact that cadmium sulfide is widely used in various fields, such as optoelectronics, photodetectors, photoresistors, etc.

Keywords: *topological index, cadmium sulfide, chemoinformatics, graph theory, molecular descriptor, randić index, zagreb index.*

Citation: A. G. Maksimov, A. D. Zavalishin, M. V. Abramov, and A. L. Tulupyev, "Chemoinformatics: Applications of Computer Science in the Analysis of Chemical Structures (on the Example of Cadmium Sulfide)," *Computer tools in education*, no. 4, pp. 44–55, 2019 (in Russian); [doi:10.32603/2071-2340-2019-4-44-54](https://doi.org/10.32603/2071-2340-2019-4-44-54)

References

1. W. Gao, A. Q. Baig, W. Khalid, and M. R. Farahani, "Molecular description of copper (II) oxide," *Macedonian Journal of Chemistry and Chemical Engineering*, vol. 36, no. 1, pp. 93–99, 2017; [doi: 10.20450/mjccce.2017.1138](https://doi.org/10.20450/mjccce.2017.1138)
2. A. Mohajeri, M. Alipour, and M. B. Ahmadi, "A graph theory study on $(ZnS)_n$ ($n = 3-10$) nanoclusters," *Chemical Physics Letters*, vol. 503, no. 1–3, pp. 162–166, 2011; [doi: 10.1016/j.cplett.2010.12.067](https://doi.org/10.1016/j.cplett.2010.12.067)
3. A. A. Taherpour, "Quantitative structural relationship and theoretical study of electrochemical properties of $C_{60}@[SWCN(5,5) - Armchair - C_nH_{20}]$ complexes," *Chemical Physics Letters*, vol. 469, no. 1–3, pp. 135–139, 2009; [doi: 10.1016/j.cplett.2008.12.039](https://doi.org/10.1016/j.cplett.2008.12.039)
4. W. Gao, M. Imran, M. K. Siddiqui, M. Naeem, and F. Jamil, "Molecular description of copper (I) oxide and copper (II) oxide," *Química Nova*, vol. 41, no. 8, pp. 874–879, 2018; [doi: 10.21577/0100-4042.20170257](https://doi.org/10.21577/0100-4042.20170257)
5. F. Deng, X. Zhang, M. Alaeiyan, A. Mehboob, and M. R. Farahani, "Topological Indices of the Pent-Heptagonal Nanosheets VC_5C_7 and HC_5C_7 ," *Advances in Materials Science and Engineering*, 2019; [doi: 10.1155/2019/9594549](https://doi.org/10.1155/2019/9594549)
6. International Agency for Research on Cancer, "Cadmium and cadmium compounds," *Monographs on evaluation of carcinogenic risks to humans*, vol. 58, pp. 119–237, 1993.

7. K. L. Chopra and S. R. Das, *Thin film solar cells*, M. M. Koltun ed., Moscow.: Mir, 1986 (in Russian).
8. I. A. Petukhov, D. A. Zuev, A. V. Shorokhova, L. S. Parshina, O. A. Novodvorskii, O. D. Khramova, A. A. Lotin, F. N. Putilin, V. F. Kozlovskii, V. K. Ivanov, M. N. Romyantseva, and A. M. Gaskov, "Cadmium sulphide thin films for photovoltaic applications," *Comp. nanotechnol.*, no. 1, pp. 68–73, 2014 (in Russian).
9. K. Paul, *Handbook of Infrared Optical Materials*, Boca Raton, FL, USA: CRC Press, 1991.
10. B. K. Mohammed, "Insight into the origin of magnetism in Iron-doped cadmium sulfide thin films from first principles calculations," *Solid State Communications*, vol. 253, no. 10–13, 2017; doi: 10.1016/j.ssc.2017.01.021
11. M. Randić, "On characterization of molecular branching," *J. Amer. Chem. Soc.*, vol. 97, no. 23, pp. 6609–6615, 1975; doi: 10.1021/ja00856a001
12. D. Amic, D. Beslo, B. Lucić, S. Nikolić, and N. Trinajstić "The vertex-connectivity index revisited," *J. Chem. Inf. Comput. Sci.*, vol. 38, no. 5, pp. 819–822, 1998; doi: 10.1021/ci980039b
13. B. Bollobas and P. Erdos, "Graphs of extremal weights," *Ars Combinatoria*, vol. 50, pp. 225–233, 1998.
14. I. Gutman and K. C. Das, "The first Zagreb index 30 years after," *MATCH Commun. Math. Comput. Chem*, vol. 50, pp. 83–92, 2004.
15. I. Gutman and N. Trinajstić, "Graph theory and molecular orbitals. Total ϕ -electron energy of alternant hydrocarbons," *Chemical Physics Letters*, vol. 17, no. 4, pp. 535–538, 1972; 10.1016/0009-2614(72)85099-1
16. E. Estrada, L. Torres, L. Rodriguez, and I. Gutman, "An at-om-bond connectivity index: Modeling the enthalpy of formation of alkanes," *Indian J. Chem.*, vol. 37A, no. 10, pp. 849–855, 1998.
17. A. Ghorbani and M. A. Hosseinzadeh, "Computing ABC4 index of nanostar Dendrimers," *Optoelectronics and Ad-vanced Materials–Rapid Communications*, vol. 4, pp. 1419–1422, 2010.
18. D. Vukičević and B. Furtula, "Topological index based on the ratios of geometrical and arithmetical means of end-vertex degrees of edges," *J. Math. Chem.*, vol. 46, pp. 1369–1376, 2009; doi: 10.1007/s10910-009-9520-x
19. A. Graovac, M. Ghorbani, and M. A. Hosseinzadeh, "Computing fifth geometric-arithmetic index for nanostar Dendrimers," *Journal of Mathematical Nanoscience*, vol. 1, no. 1-2, pp. 33–42, 2011.
20. M. V. Abramov, T. V. Tulupyeva, and A. L. Tulupjev, "Sotsioinzhenernye ataki: sotsial'nye seti i otsenki zashchishchennosti pol'zovatelei"[Socioengineering attacks: social networks and user security ratings], Saint Petersburg, Russia: GUAP, 2018.
21. A. A. Azarov, T. V. Tulupyeva, A. V. Suvorova, A. L. Tulupjev, M. V. Abramov, and R. M. Yusupov, "Sotsioinzhenernye ataki. Problemy analiza"[Socioengineering attacks. Analysis of the problem], Saint Petersburg, Russia: Nauka, 2016.
22. A. O. Khlobystova, M. V. Abramov, A. L. Tulupjev, and A. A. Zolotin, "Search for the shortest trajectory of a social engineering attack between a pair of users in a graph with transition probabilities," *Information and Control Systems*, no. 6, pp. 74–81, 2018 (in Russian).
23. A. V. Suvorova, K. R. Smirnova, E. A. Budin, T. V. Tulupyeva, A. L. Tulupjev, and M. V. Abramov, "Research Project as a Tool for Teaching Text Analysis Methods: Predicting the Post Class in the Social Network," *Computer tools in education*, no. 3, pp. 49–64, 2018 (in Russian).
24. N. A. Kharitonov, A. G. Maximov, and A. L. Tulupjev, "Algebraic Bayesian Networks: Naïve Frequentist Approach to Local Machine Learning Based on Imperfect Information from Social Media and Expert Estimates," S. Kuznetsov, A. Panov, eds, in *Proc. RCAI 2019: Artificial Intelligence. Communications in Computer and Information Science*, vol. 1093, 2019, pp. 234–244; doi: 10.1007/978-3-030-30763-9_20

Received 05.11.2019, The final version — 20.12.2019.

Anatolii G. Maksimov, junior researcher, Laboratory of Theoretical and Interdisciplinary Problems of Informatics, SPIIRAS; student, Computer Science Department, SPbU, agm@dscs.pro

Arsenii D. Zavalishin, junior researcher, Laboratory of Theoretical and Interdisciplinary Problems of Informatics, SPIIRAS; student, Computer Science Department, SPbU, adz@dscs.pro

Maxim V. Abramov, PhD, senior researcher, Laboratory of Theoretical and Interdisciplinary Problems of Informatics, SPIIRAS; Associate Professor, Computer Science Department, SPbU, mva@dscs.pro

Alexander L. Tulupjev, PhD, Dc. Sci., professor, Computer Science Department, SPbU; Principal Researcher, Laboratory of Theoretical and Interdisciplinary Problems of Informatics, SPIIRAS, alt@dscs.pro