

ПОСТРОЕНИЕ ВСЕХ НЕИЗОМОРФНЫХ ХИМИЧЕСКИХ ГРАФОВ  
ИЗ ЗАДАННОГО НАБОРА СТРУКТУРНЫХ ФРАГМЕНТОВ

С.Г.Молодцов, В.Н.Пиоттух-Пелецкий

Проблема генерации структур химических соединений из заданного набора структурных фрагментов представляет большой интерес с точки зрения химических исследований. За последние 15 лет появился ряд работ, посвященных разработке алгоритмов генерации графов (в первом приближении удовлетворительно представляющих структуры химических соединений) и их программной реализации [1-6]. Общим недостатком большинства этих работ является большое время, необходимое для решения сравнительно простых задач генерации графов. Одной из причин их низкой эффективности, по-видимому, является генерация изоморфных графов, которые приходится впоследствии выявлять при помощи достаточно трудоемкой процедуры проверки на изоморфизм.

Целью настоящей работы являются создание и программная реализация алгоритма генерации только неизоморфных графов с помеченными вершинами и кратными ребрами для последующего использования при построении исчерпывающих наборов изомеров химических соединений заданного состава. Основными областями применения данной разработки в химии могут быть задачи установления строения химических соединений по заданному набору спектральных характеристик и построение структурных формул органических соединений, гипотетически обладающих заданным видом биологической активности или какого-то другого свойства. В обоих случаях химик имеет информацию о фрагментах структуры, являющихся строительным материалом для генерации графов, и нуждается в построении исчерпывающего набора структурных формул, отвечающих заданному фрагментарному составу и ограничениям, задаваемым исследователем. Особое внимание в данной

работе было уделено средствам обеспечения эффективности взаимодействия пользователя с программой генерации. Эти средства, во-первых, предназначены для сокращения времени генерации семейств графов, и, во-вторых, предоставляют пользователю большие возможности для задания ограничений на генерируемые структуры. Оба этих пункта взаимосвязаны, поскольку включение модулей проверки ограничений в алгоритм генерации позволяет получить наибольшую экономию времени.

## 1. Алгоритм генерации

Предлагаемый алгоритм генерации неизоморфных мультиграфов с заданными наборами меток и степеней вершин основан на методе ветвей и границ. Аналогичная схема была применена в [4] для генерации всех неизоморфных обыкновенных графов с заданным распределением степеней вершин. Основная идея состоит в том, что в каждом множестве изоморфных между собой графов выбирается по некоторому критерию один граф, называемый каноническим, и в процессе генерации стремятся строить только канонические графы. Преимуществом такого подхода является отсутствие необходимости в проверке изоморфизма с ранее построенными графами. Приведенные в [8] результаты работы программы показали довольно высокую эффективность алгоритма при построении графов с заданным распределением степеней вершин, в частности, однородных графов. Но в различных приложениях, например, при химических исследованиях, требуется генерировать графы с помеченными вершинами и кратными ребрами (связями). С этой целью алгоритм [4] был переработан и дополнен.

Поставим в соответствие мультиграфу (неориентированному, без петель, с кратными ребрами)  $G = (V, E)$  со множеством вершин  $V$ ,  $|V| = n$ , и множеством ребер  $E \subset V \times V$  матрицу смежности  $A = (a_{ij})$ , где  $a_{ij}$  равно кратности связи между  $i$ -й и  $j$ -й вершинами, в частности,  $a_{ij} = 0$ , если эти вершины несмежны. Очевидно, что представление графа в виде матрицы смежности неоднозначно, оно зависит от нумерации вершин. Пусть вершины графа помечены, причем одинаково помеченные вершины имеют одинаковую степень. Упорядочим метки каким-нибудь образом, например, по количеству вершин с данной меткой, и будем рассматривать только такие нумерации вершин графа, называемые допустимыми, которые отвечают этому порядку, т.е. вершины с меньшей меткой имеют меньший номер. Таким образом, мно -

множество  $\Pi = \{1, 2, \dots, n\}$  разбивается на ряд подмножеств  $M_1, M_2, \dots, M_k$ , где  $M_i$  — множество номеров вершин с  $i$ -й меткой, причем  $\bigcup_{i=1}^k M_i = \Pi$ ,  $M_i \cap M_j = \emptyset$  при  $i \neq j$  и  $u < v$ , если  $u \in M_i$ ,  $v \in M_j$ ,  $i < j$ .

Введем на множестве матриц смежности одинакового размера отношение линейного порядка следующим образом: поставим в соответствие матрице  $A = (a_{ij})$   $n^2$ -мерный вектор  $W(A) = (a_{11}, a_{12}, \dots, a_{1n}, \dots, a_{nn})$ , т.е. упорядочим элементы матрицы по строкам. Тогда определим  $(a_{ij}) = A < B = (b_{ij}) \Leftrightarrow W(A) < W(B)$ , где порядок на множестве  $n^2$ -мерных векторов определен лексикографически.

Пусть  $S(M_i)$  — симметрическая группа подстановок, действие которой на матрицу смежности  $A$  приводит к одинаковой перестановке строк и столбцов матрицы с номерами из  $M_i$ , т.е. к перенумерации вершин графа с  $i$ -й меткой. Тогда действие группы  $S = S(M_1, M_2, \dots, M_k) = S(M_1) \oplus S(M_2) \oplus \dots \oplus S(M_k)$  приводит к допустимой перенумерации вершин графа.

**ОПРЕДЕЛЕНИЕ 1.** Графы, матрицы смежности которых переводятся друг в друга в результате перестановки из  $S$ , называются изоморфными.

**ОПРЕДЕЛЕНИЕ 2.** Матрицу смежности  $A = (a_{ij})$  графа с помеченными вершинами назовем канонической, если она не увеличивается под действием любой подстановки из группы  $S$ .

Очевидно, что две различные канонические матрицы смежности соответствуют неизоморфным графам.

Построение канонических матриц смежности графов с помеченными вершинами отличается от алгоритма [4] введением возможности образования кратных ребер и заданием максимальных кратностей связей между вершинами с разными или одинаковыми метками, в частности, запретом связей.

Назовем сильно-канонической матрицей смежности такую частично заполненную матрицу, которая не увеличивается под действием подстановок из  $S$ , переставляющих заполненные и незаполненные строки. На каждом шаге предлагаемого алгоритма строится некоторое сильно-каноническое пополнение частично-заполненной матрицы смежности, полученной на предыдущем шаге. Процедура построения состоит из следующих пунктов.

I. Заполняется первая незаполненная строка очередным способом с учетом максимальных кратностей связей между вершинами. Если

все способы заполнения исчерпаны, то происходит возврат на предыдущий шаг.

2. Проверяется допустимость и возможность форсированного заполнения еще не заполненных строк. Незаполненная строка является допустимой, если ее можно заполнить таким образом с учетом максимальных кратностей связей между вершинами, что сумма ненулевых элементов будет равна степени соответствующей вершины. Форсированное заполнение происходит в случае, когда заполнение возможно только одним способом. Если полученная матрица недопустима, то переходим к п. I для рассмотрения других пополнений.

3. Проверяются условия связности и сильной каноничности и если они не выполнены, переходим к п. I. Если полученная матрица заполнена полностью, то она каноническая, иначе переходим на следующий шаг для дальнейшего пополнения частично-заполненной матрицы смежности.

При этом алгоритм проверки сильной каноничности был модифицирован по сравнению с описанным в [7]: на каждом шаге форсированно заполненные строки полностью определяются тем, как была заполнена первая незаполненная строка (пусть ее номер будет  $m$ ). Тогда при поиске подстановки из  $S$ , действие которой увеличивает частично-заполненную матрицу смежности  $A = (a_{ij})$ , достаточно проверять на сильную каноничность только головной фрагмент  $A_m$ , состоящий из первых  $m$  строк матрицы  $A$ , даже если следующие за  $m$ -й строкой строки были заполнены форсированно. Преимуществом такого подхода является существенное сокращение дерева перебора подстановок из  $S$ , когда многие строки заполняются форсированно. В частности, все строки, стоящие в конце матрицы смежности и соответствующие одновалентным вершинам, заполняются форсированно.

Во многих случаях имеется необходимость наложения структурных ограничений на генерируемые графы. Для этого в алгоритме предусмотрена возможность задания обязательных и запрещенных фрагментов (в том числе и пересекающихся) в генерируемых структурах. При этом вершины в фрагментах как обязательных, так и запрещенных могут иметь универсальную метку, т.е. они могут отождествляться с любой меткой, что позволяет задавать, например, ограничения на размеры циклов.

## 2. Особенности программной реализации алгоритма

Программа, реализующая описанный выше алгоритм, написана на языке ФОРТРАН ЕС ЭЕМ. Программа генерирует все связанные неизоморфные графы с помеченными вершинами и (возможно) кратными ребрами, удовлетворяющие заданному структурному ограничению в виде обязательных и запрещаемых фрагментов. Результаты работы программы — матрицы смежности сгенерированных графов могут быть выданы пользователю или сохранены в упакованном виде для целей дальнейшей обработки. По желанию пользователя возможна программная разрисовка сгенерированных структур на АЦПУ.

Генерируемые структуры определяются прежде всего множеством вершин: числом типов вершин, количеством вершин в каждом типе и их степенями. В частности, таким множеством вершин может служить молекулярная формула химического соединения. Предполагается, что порядок задания типов определяет их относительное старшинство. Например, в случае генерации всех изомеров бензола состава  $C_6H_6$ , используя в качестве вершин отдельные атомы, множество вершин задается следующим образом:

6 вершин C, имеющих степень 4,  
6 вершин H, имеющих степень 1.

Множество вершин генерируемых структур может соответствовать не только атомам, но и более крупным структурным фрагментам, удовлетворяющим следующим условиям:

- фрагмент имеет эквивалентные свободные ребра;
- при связывании некоторых свободных ребер фрагмента оставшиеся свободные ребра должны сохранять эквивалентность;
- набор фрагментов должен быть внутренне независимым, т.е. ни один из фрагментов набора не должен получаться при связывании других фрагментов между собой.

Описанным условиям удовлетворяют фрагменты, имеющие свободные ребра только у одной вершины, например,  $\geq CH$ ,  $>CH_2$ ,  $-CH_3$ ,  $>C=O$ ,  $-COOH$ . Использование фрагментов структуры в тех случаях, когда они могут быть заранее известны пользователю наряду с молекулярной формулой, в значительной степени сокращает как множество генерируемых графов, так и время генерации. Например, при генерации всех изомеров с молекулярной формулой  $C_8H_{16}$  было сгенерировано 73 графа, не содержащих кратных ребер, за 11 сек. В случае задания в качестве вершин фрагментов (2 вершины типа  $\geq CH$ , 4 вер-

шины  $>CH_2$ , 2 вершины  $-CH_3$ ), суммарный элементный состав которых также  $C_6H_{16}$ , было сгенерировано 18 графов, не содержащих кратных ребер, за 2 секунды.

В ряде случаев возникает необходимость наложения ограничений на образование или кратность связей между определенными типами вершин. Ограничения этого вида задаются с помощью матрицы максимальных кратностей связей между вершинами отдельных типов. Запрет образования связи между вершинами определенных типов указывается с помощью нулевого значения соответствующего элемента матрицы. Например, для построения всех изомеров с молекулярной формулой  $C_6H_8O_2$  без тройных связей и без связей O-O задаются:

6 вершин C, имеющих степень 4,  
 2 вершины O, имеющих степень 2,  
 8 вершин H, имеющих степень 1.

Матрица максимальных кратностей связей в этом случае выглядит следующим образом:

$$\begin{matrix} C & \begin{pmatrix} 2 & 2 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} \\ O & \\ H & \end{matrix}$$

Запрещаются тройные связи, а также связи O-O и H-H (последние привели бы к образованию несвязных графов).

Часто возникает потребность построения графов, удовлетворяющих некоторым более сложным ограничениям на содержащиеся в них фрагменты. Такие ограничения могут быть как в отношении запрета на вложение в генерируемые структуры определенных (достаточно сложных) фрагментов, так и в отношении обязательного присутствия заданных фрагментов (предполагается, что необходимые или запрещаемые фрагменты могут быть построены из вершин, используемых при генерации). Фрагменты, используемые в качестве ограничений, могут задаваться двумя способами: с возможностью пересечений друг с другом и без возможности пересечений. Например, мы можем потребовать, чтобы все генерируемые графы содержали как минимум два фрагмента (рис.1). Если мы допускаем возможность пересечения фрагментов, то нас удовлетворят и структуры, изображенные на рис.2.

В случае запрета на пересечения графы, содержащие данные фрагменты, будут отбракованы.

При описании структуры фрагментов ограничений имеется возможность задания для некоторых (или для всех) вершин "универсаль-

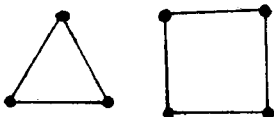


Рис. 1

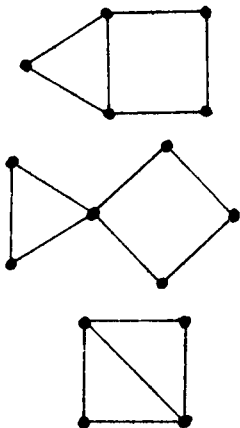


Рис. 2

ного типа" - XX. Считается, что такая вершина может соответствовать вершине любого типа в генерируемых графах. Использование этого средства позволяет накладывать ограничения на характеристики циклической системы генерируемых графов независимо от типов составляющих вершин.

Разработанная программа была проверена на ряде известных задач построения изомеров [6,9] и графов [8] и показала высокую эффективность.

В работе [6] приводятся в основном результаты по генерации ациклических графов (деревьев); в тех же случаях, когда генерировались циклические структуры, время генерации было довольно большим: построение всех 98 изомеров с молекулярной формулой  $C_4H_7NO$ , содержащих трехчленный цикл  $\widehat{C-N-C}$ , требовало как минимум 2 минуты (быстродействие ЭЕМ 100 тыс. оп/сек). Генерация по нашей программе заняла 20 секунд машинного времени на ЕС ЭЕМ (быстродействие 180 тыс.оп/сек).

С помощью разработанной программы были получены полные списки связанных однородных мультиграфов степени 3. Число сгенерированных графов и время их генерации приводятся в таблице. Заметим, что

Т а б л и ц а

Число вершин	Число графов	Время генерации
6	6	0,4 сек
8	20	1 сек
10	91	12 сек
12	509	90 сек
14	3608	850 сек

построение всех 509 неизоморфных простых (без кратных ребер) связанных однородных графов степени 3 потребовало 2,7 минуты машинного времени (в [8] - 3,3 минуты при быстродействии 300 тыс.оп/сек).

Изомеров с молекулярной  $C_5H_{10}O_2$  оказалось ровно 400, время генерации 20 сек, а изомеров с молекулярной формулой  $C_4H_9NO_2$  - 1341, время генерации 1 минута. Отметим, что такие примеры в [9]

не были приведены, по-видимому, их решение потребовало бы уже слишком много машинного времени.

Из приведенных примеров видно, что предложенный алгоритм об- ладает по сравнению с другими большей универсальностью и меньшей вычислительной сложностью, позволяет ставить и решать задачи, ре- шение которых до этого было либо невозможным, либо требовало больш- их затрат машинного времени. Предлагаемая реализация алгоритма включает в себя ряд средств, позволяющих ограничивать вид генери- руемых графов с учетом требования пользователей. Кроме того, генери- руемые графы могут быть изображены на алфавитно-цифровом печа- тающем устройстве с целью повышения наглядности результатов.

#### Л и т е р а т у р а

1. НЕХОРОШЕВ С.А., КОПТУГ В.А. Построение полного набора воз- можных структурных формул исследуемого соединения по заданному набору фрагментов с помощью ЭВМ (усовершенствованный вариант). - Изв. СО АН СССР, сер.хим.наук, 1975, вып.6, №14, с. 66-70.
2. BROWN H., MASINTER I. An algorithm for the construction of the graphs of organic molecules.-Discrete Math.,1974,N8,p.227.
3. BROWN H. Molecular structure elucidation.-SIAM J.Appl. Math.,1977,v.32,N 3,p.534-551.
4. ФАРАДЖЕВ И.А. Генерирование неизоморфных графов с задан - ным распределением степеней вершин. -В кн.: Алгоритмические ис- следования в комбинаторике. М., 1978, с. 11-19.
5. СЕРОВ В.В., ЭЛЯШБЕРГ М.Е., ГРИБОВ Л.А. Система математи- ческого синтеза и анализа всех структурных формул органических соединений, соответствующих заданной брутто-формуле.-Докл.АН СССР, 1975, т. 224, с. 109-111.
6. МИТРОФАНОВ Ю.П., РАЗНИКОВ В.В., ШКУРОВ В.А. Метод автома- тического построения топологических структурных формул органиче- ских соединений по заданной брутто-формуле.-Ж.анал.химии, 1982, т. 37, вып. 8, с. 1477-1483.
7. Алгоритм проверки каноничности частично-заполненной мат- рицы смежности графа /Зайченко В.А., Иванов А.В., Розенфельд М.З. и др. -В кн.: Алгоритмические исследования в комбинаторике. М., 1978, с. 19-25.
8. БАРАЕВ А.М., ФАРАДЖЕВ И.А. Построение и исследование на ЭВМ однородных и однородных двудольных графов. -В кн.: Алгоритми- ческие исследования в комбинаторике. М., 1978, с. 25-28.
9. ЭЛЯШБЕРГ М.Е., ГРИБОВ Л.А., СЕРОВ В.В. Молекулярный спект- ральный анализ и ЭВМ. -М.: Наука, 1980. - 163 с.

Поступила в ред.-изд.отд.  
30 марта 1984 года