Записки научных
семинаров ПОМИ
Том Kol, 2019 г.

А. Э. Гутерман, Е. М. Крейнес, Н. В. Остроухова

2-СЛОВА: ИХ ГРАФЫ И МАТРИЦЫ

§1. Введение

2-слова и соответствующие им простые сборные графы являются важными понятиями нетривиально связанных между собой разделов алгебры и геометрии. Следующие два направления исследований активно развиваются в связи с различными приложениями этих математических понятий. Первое из них – изучение численных инвариантов 2-слов, таких как сборное число, минимальное реализующее число, род и т.д. Второе связано с конструкциями, позволяющими строить новые семейства сборных графов по имеющимся.

В данной работе мы рассматриваем конечные графы $\Gamma = (V, E)$, с множеством вершин V и множеством ребер $E \subseteq V \times V$. Граф может содержать петли и кратные ребра.

Определение 1.1. *Степень* или *валентность* вершины $v \in V$ – это число ребер, инцидентных данной вершине. Если ребро дважды инцидентно вершине (такое ребро называют петлей), оно учитывается дважды.

Определение 1.2 ([1, с. 3022])**.** Циклический порядок для кортежа из k элементов $(x_1, x_2, x_3, \dots, x_{k-1}, x_k)$ – это множество

$$(x_1, x_2, x_3, \dots, x_{k-1}, x_k)^{cyc} = \{(x_1, x_2, x_3, \dots, x_{k-1}, x_k),$$
$$(x_2, x_3, \dots, x_{k-1}, x_k, x_1), (x_3, \dots, x_{k-1}, x_k, x_1, x_2), \dots,$$
$$(x_k, x_1, x_2, x_3, \dots, x_{k-1}), (x_k, x_{k-1}, x_{k-2}, \dots, x_2, x_1),$$
$$(x_{k-1}, x_{k-2}, \dots, x_2, x_1, x_k), (x_{k-2}, \dots, x_2, x_1, x_k, x_{k-1}), \dots,$$
$$(x_1, x_k, x_{k-1}, x_{k-2}, \dots, x_2)\},$$

т.е. все циклические сдвиги данного кортежа и все циклические сдвиги этого же кортежа, записанного в обратном порядке.

Ключевые слова: сборные графы, 2-слова, матрицы индицентности.
Работа выполнена при поддержке гранта РНФ 17-11-01124.

Чтобы задать циклический порядок, достаточно одного элемента множества $(x_1, x_2, x_3, \ldots, x_{k-1}, x_k)^{cyc}$, остальные получаются циклическими сдвигами и записыванием их в обратном порядке.

Определение 1.3. Если циклический порядок ребер, инцидентных вершине v зафиксирован, говорят, что вершина *упорядоченная* (или, иногда, *регулярная*).

Замечание 1.4. 1. Например, для графа, вложенного в ориентированную поверхность, все его вершины регулярные.

2. Для каждого из ребер упорядоченной вершины корректно определены его *соседи*.

Регулярная вершина валентности n называется также n-регулярной.

Пример 1.5. На рисунке 1 изображена упорядоченная вершина валентности 4 с циклическим порядком ребер (e_1, e_2, e_3, e_4). Нетрудно видеть, что ребра e_2 и e_4 являются соседями ребра e_1 (или e_3) в вершине v.

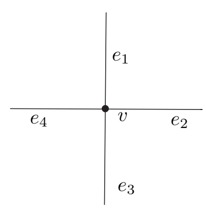


Рис. 1. Упорядоченная вершина валентности 4

В работе [5] исследуется особый тип упорядоченных графов, называемых сборными графами. Эти графы встречаются в генетике и используются для описания эпигенетических геномных перестроек.

Определение 1.6. *Сборным графом* называется конечный связный граф, в котором все вершины упорядоченные и имеют валентность 1 или 4.

Определение 1.7. Вершины валентности 1 называются *концевыми*.

Определение 1.8. Число 4-регулярных вершин сборного графа Γ называется *порядком* Γ и обозначается $|\Gamma|$.

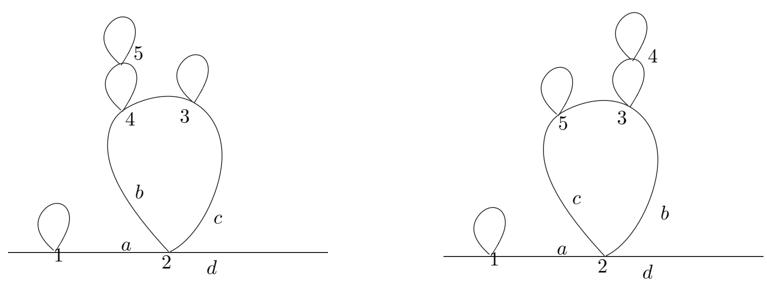
Определение 1.9. Сборный граф называется *тривиальным*, если $|\Gamma| = 0$.

Определение 1.10. Говорят, что сборные графы $\Gamma_1 = (V_1, E_1)$, $\Gamma_2 = (V_2, E_2)$ *изоморфны*, если $|\Gamma_1| = |\Gamma_2|$ и существует изоморфизм $\phi : V_1 \to V_2$, такой, что:

1. для любых $u, v \in V_1$ ребро $(u, v) \in E_1$ тогда и только тогда, когда $(\phi(u), \phi(v)) \in E_2$;
2. для любой $u \in V_1$ циклический порядок ребер в u совпадает с циклическим порядком их ϕ-образов в $\phi(u)$.

Заметим, что два графа могут быть изоморфны как абстрактные графы, но не изоморфны как сборные графы.

Пример 1.11. Рассмотрим графы Γ_A и Γ_B на рисунке 2, порядок ребер и отношение соседства очевидно из рисунка. К примеру, порядок ребер графа Γ_A в вершине 2 определяется как $(a, b, c, d)^{cyc}$, в то время как порядок ребер графа Γ_B в вершине 2 соответствует $(a, c, b, d)^{cyc}$. Эти графы, очевидно, изоморфны в общем смысле. Тем не менее, этот изоморфизм не сохраняет циклический порядок ребер в вершине 2. В графе Γ_A ребра d и b – соседи ребра a в вершине 2, но в графе Γ_B ребро a является соседом для c и d в вершине 2.



(A) Граф Γ_A (B) Граф Γ_B

Рис. 2. Сборные графы, изоморфные в общем смысле, но не как сборные графы.

Определение 1.12. *Путь* из вершины u в вершину v графа Γ – это последовательность вершин $u, w_1, \ldots, w_l, v \in V(\Gamma)$ и ребер (u, w_1), $(w_1, w_2), \ldots, (w_l, v) \in E(\Gamma)$, в которой вершины и ребра не обязательно различны. Путь, в котором вершины не повторяются, называется *простым.* Путь, состоящий всего из одной вершины, называется *одноточечным.*

Определение 1.13. *Трансверсальный путь*, или *трансверсаль* – это путь в графе Γ, в котором все ребра попарно различны, а последовательные ребра не являются соседями в их общей вершине в смысле введенного выше упорядочения вершин.

Определение 1.14. Трансверсаль называется *эйлеровой*, если она проходит по всем ребрам графа Γ.

Определение 1.15. Сборный граф, в котором существует эйлерова трансверсаль, называется *простым сборным графом.*

Определение 1.16. Путь, в котором все вершины попарно различны, а последовательные ребра являются соседями в их общей вершине, называется *полигональным.*

Лемма 1.17. *Пусть* Γ *– простой сборный граф с двумя концевыми вершинами,* $|\Gamma| = n$. *Тогда граф* Γ *состоит из* $2n+1$ *ребра. Если у графа* Γ *нет концевых вершин, то он содержит* $2n$ *ребер.*

Доказательство. Поскольку в простом сборном графе имеется эйлерова трансверсаль, вершин валентности 1 либо нет, либо две, а все остальные вершины имеют валентность 4.

Посчитаем число ребер. Для этого просуммируем валентности всех вершин и разделим результат на 2, поскольку каждое ребро учтено в этой сумме дважды. В случае двух концевых вершин число ребер равно $\frac{4\cdot n+2}{2} = 2n+1$, а в случае, когда концевых вершин нет, число ребер равно $\frac{4\cdot n}{2} = 2n$. □

Если в сборном графе Γ, $|\Gamma| = n$, существует эйлеров трансверсальный путь, то можно выбрать ориентацию этого пути и получить таким образом ориентированный (или направленный) простой сборный граф. Таким образом, концевых вершин в простом сборном графе или ноль, или две. В случае двух концевых вершин ориентированная эйлерова трансверсаль начинается в одной из них (обозначим эту вершину 0), а заканчивается в другой – обозначим эту вершину $n+1$.

Определение 1.18. Два трансверсальных пути эквивалентны, если они совпадают или один из них – это другой путь, прйденный в обратном порядке.

Далее, если иное не указано явным образом, все графы считаются простыми сборными графами с ровно двумя концевыми вершинами.

Сборные графы естественным образом связаны со специальным классом слов.

Определение 1.19. Сборное слово или 2-*слово* – это слово в некотором алфавите $S = \{a_1, a_2, \dots\}$ такое, что каждая буква a_i либо содержится в слове ровно два раза, либо не содержится вовсе.

Определение 1.20. Слово $w^R = a_{i_k} \dots a_{i_1}$ называется обратным к слову $w = a_{i_1} \dots a_{i_k}$.

Определение 1.21. 2-слова называются *эквивалентными*, если после переименования некоторых букв они либо совпадают, либо являются обратными друг для друга.

Пример 1.22. Слово $w = 123321$ эквивалено своему обратному, а слово $w' = 213132$ совпадает с w^R, если переобозначить буквы $1 \leftrightarrow 3$.

Пустое 2-слово обозначается ϵ.

Следующая теорема связывает сборные графы и 2-слова.

Теорема 1.23 ([1, лемма 3.8]). *Классы эквивалентности 2-слов находятся в биективном соответствии с классами изоморфизма простых сборных графов.*

Пример ниже иллюстрирует данную теорему.

Пример 1.24. Сборные 2-слова для графов Γ_A и Γ_B из примера 1.11, изоморфных как абстрактные графы, но не как сборные графы, – это 1123345542 и 1123443552 соответственно. Легко проверить, что эти сборные слова не эквивалентны.

Для построения слова по графу можно использовать следующий алгоритм.

Алгоритм 1.25.

(1) *Пометим все 4-валентные вершины графа Γ натуральными числами $1, \dots, |\Gamma|$.*
(2) *Выберем одну из вершин валентности 1 в качестве начала трансверсали и назовем ее начальной.*
(3) *Будем двигаться по трансверсальному пути, стартуя, с начальной вершины, и будем записывать номер 4-валентной вершины каждый раз, когда трансверсаль через нее проходит.*

Поскольку валентность каждой из помеченных вершин равна 4, а трансверсальный путь, по определению, проходит по всем ребрам, он пройдет через каждую из 4-валентных вершин дважды, таким образом мы получим 2-слово.

Предположим, нам дано 2-слово $\omega = w_1 w_2 \dots w_{2n-1} w_{2n}$. В этом случае w_i обозначает i-ую букву в слове и, поэтому, не все w_i различны. Для того, чтобы нарисовать соответствующий граф, можно воспользоваться следующим алгоритмом.

Алгоритм 1.26.

(1) *Нарисуем начальную вершину и исходящее из нее ребро.*

(2) *Поставим на этом ребре вершину и пометим ее буквой w_1, нарисуем ребро, выходящее из вершины w_1. Отметим, что вершина w_1 пока получилась 2-валентной.*

(3) *Каждая из букв w_i, $i \in \{2, 3, \dots, 2n-1\}$ обрабатывается следующим образом.*

   (a) *На предыдущем шаге построен граф, выделена вершина, w_{i-1}, и исходящее из нее ребро, второй конец которого пока не определен.*

   (b) *В случае, если буква w_i ранее не встречалась в слове ω, поставим новую вершину на ребре, исходящем из w_{i-1}, пометим ее буквой w_i, нарисуем исходящее ребро и перейдем к началу шага 3. Отметим, что вершина w_i получилась 2-валентной, что означает, в частности, что в слове ω есть еще буквы, кроме w_i.*

   (c) *Если буква w_i ранее встречалась в ω, то, по построению, граф содержит 2-валентную вершину v_k, помеченную буквой w_i. Обозначим ребра, инцидентные v_k, через e_1^k, e_2^k. В этом случае, мы соединяем вершину w_{i-1} с вершиной v_k таким образом, чтобы новое ребро e_3^k входило в вершину между ребрами e_1^k и e_2^k. Нарисуем исходящее ребро e_4^k между ребрами e_1^k и e_2^k там, где между ними пока нет ребер. Заметим, что в силу определения 1.2 неважно, получим мы в вершине k ребра в порядке $(e_1^k, e_3^k, e_2^k, e_4^k)$ или $(e_1^k, e_4^k, e_2^k, e_3^k)$ (по часовой стрелке, начиная с e_1^k), поскольку оба кортежа задают один и тот же циклический порядок ребер в вершине.*

   (d) *Если в слове ω есть еще буквы, кроме w_i, возвращаемся в начало шага 3, иначе переходим к следующему шагу.*

(4) *Когда мы обработали последнюю букву слова, очевидно, это ее второе вхождение, мы завершаем нарисованный «перекресток» ребер конечной вершиной.*

На рисунке 3 изображен процесс построения сборного графа по слову 1221.

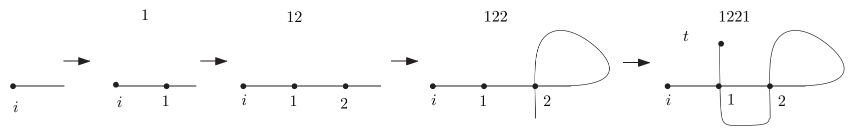


Рис. 3. Построение простого сборного графа по слову 1221.

Для заданного 2-слова w соответствующий ему сборный граф мы будем обозначать Γ_w. Пустое слово ϵ соответствует двум вершинам 0 и 1, соединенным ребром.

Определение 1.27. *Композиция* $\Gamma_1 \circ \Gamma_2$ двух ориентированных простых сборных графов Γ_1 и Γ_2 – это граф, который получается если отождествить конечную вершину Γ_1 и начальную вершину Γ_2, после чего забыть об этой вершине.

Замечание 1.28. Очевидно, что композиция простых сборных графов – простой сборный граф.

Композиция двух слов – это их конкатенация. Из определения немедленно следует, что для сборных слов w_1, w_2 мы имеем $\Gamma_{w_1} \circ \Gamma_{w_2} = \Gamma_{w_1 w_2}$.

Вообще говоря, графы $\Gamma_1 \circ \Gamma_2$ и $\Gamma_2 \circ \Gamma_1$ не изоморфны. Рассмотрим, к примеру, Γ_{aa} и Γ_{bbcddc}. Имеем: $\Gamma_1 \circ \Gamma_2 = \Gamma_{aabbcddc}$, в то время как $\Gamma_2 \circ \Gamma_1 = \Gamma_{bbcddcaa}$.

Определение 1.29. Композиция $\underbrace{\Gamma \circ \Gamma \circ \cdots \circ \Gamma}_{k}$ называется k - той степенью графа Γ и обозначается Γ^k.

Введем теперь понятия гамильтонова множества полигональных путей и, соответственно, сборного числа, которые являются главным предметом наших исследований. Эти характеристики сборных графов подробно рассматривались в [1, 2, 5].

Определение 1.30. Два пути не пересекаются, если у них нет общих вершин.

Мы рассматриваем полигональные непересекающиеся пути, покрывающие все вершины сборного графа.

Определение 1.31. Множество $\{\gamma_1, \gamma_2, \ldots, \gamma_k\}$ попарно непересекающихся полигональных путей в графе Γ называется гамильтоновым, если их объединение содержит все 4-валентные регулярные вершины графа Γ.

Например, множество всех вершин графа $V(\Gamma)$ является гамильтоновым множеством одноточечных путей.

Определение 1.32. Полигональный путь γ называется гамильтоновым, если множество $\{\gamma\}$ является гамильтоновым.

Пусть Γ – нетривиальный сборный граф.

Определение 1.33. Сборное число Γ (обозначается $\mathrm{An}(\Gamma)$) – это минимальный размер гамильтонова множества полигональных путей, т.е. $\min\{k :$ существует гамильтоново множество полигональных путей $\{\gamma_1, \gamma_2, \ldots, \gamma_k\}$ в $\Gamma\}$.

Определение 1.34. Граф, для которого $\mathrm{An}(\Gamma) = 1$, называется реализуемым, иначе нереализуемым.

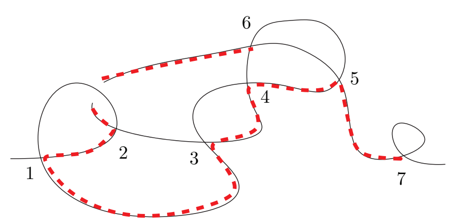


Рис. 4. Реализуемый граф со сборным словом 12134564326577 и гамильтоновым полигональным путем, обозначенным пунктиром.

Определение 1.35. Для натурального n минимальное реализующее число определяется как минимальный размер сборного графа со сборным числом n: $\mathrm{R}_{\min}(n) = \min\{|\Gamma| : \mathrm{An}(\Gamma) = n\}$.

Хорошо известно, что сборные графы используются для описания эпигенетических геномных перестроек, см., например, [6]. Следуя работе [1], мы кратко опишем здесь соответствующий процесс и некоторые попытки его моделирования. Более подробное описание можно найти в [1, 6] и упомянутых там источниках.

Существует два типа нуклеотидов, микронуклеус и макронуклеус, возможно в виде нескольких копий. Во время размножения происходит обмен только генами микронуклеуса. После коньюгации старый макронуклеус разрушается, а новый формируется с помощью одного из вновь образованных микронуклеусов. Эти процессы формирования ДНК включают удаление 95-98% сгенерированного ДНК, эффективную ликвидацию всего т.н. «мусорного» ДНК – некодирующих последовательностей (англ. internal eliminated sequences, IESs), которые прерывают закодированный ген. В силу того, что некодирующие последовательности разделяют кодирующие участки в микронуклеусе, каждый ген в макронуклеусе может быть представлен несколькими не идущими подряд сегментами (англ. macronuclear destined sequences, MDSs) микронуклеуса. Более того, для тысяч генов соответствующие этим сегментам части микронуклеуса могут быть перемешаны или идти в обратном порядке.

Известно несколько моделей рекомбинации ДНК [6–8, 10], а также гипотеза, что в процессе рекомбинации участвует дополнительная молекула [2, 10]. Последняя модель была экспериментально подтверждена, см. [9].

На основе этих наблюдений в [2] была представлена теоретическая модель, в которой молекулы в процессе рекомбинации представлены как пространственные графы. Эта модель описывает ген микронуклеуса как граф, все вершины которого регулярны и имеют валентности 1 или 4. Каждая 4-валентная вершина представляет собой участок, на котором происходит рекомбинация. Один ген микронуклеуса соответствует сборному графу, имеющему эйлеров путь, в котором последовательные ребра не являются «соседями» в их общей вершине. Заметим, что последовательность вершин, записанных в том порядке, в котором через них проходит эйлеров путь, как раз и формирует сборное слово.

Целью нашей работы является описание простых сборных графов с помощью их матриц инцидентности. В частности, мы характеризуем матрицы, соответствующие нескольким стандартными сериям графов, и переводим на матричный язык некоторые важные с точки зрения генетики действия, производимые на графах. Например, мы описываем процедуру добавления петель и конкатеницию графов.

Настоящая работа имеет следующую структуру. В §2 приведены основные свойства матриц инцидентности сборных графов. §3 содержит конкретные примеры матриц инцидентности для специальных серий сборных графов. В §§4 и 5 мы описываем на языке матриц некоторые стандартные процедуры, применяемые к сборным графам, такие как (внутренняя) петельная подстановка и композиция, используемые в генетике.

§2. Матрицы инцидентности

В этом разделе мы напомним определение матрицы инцидентности произвольного графа и опишем несколько свойств матриц инцидентности простых сборных графов.

Определение 2.1. Пусть Γ – произвольный, не обязательно простой сборный, граф с вершинами $v_1, \ldots, v_n$ и ребрами $e_1, e_2, \ldots, e_m$. Матрица инцидентности графа Γ – это целочисленная матрица $I(\Gamma) = (a_{ij})$ размера $n \times m$, определяемая следующим образом:

$$a_{ij} = \begin{cases} 1, & \text{если } v_i \text{ инцидентно } e_j, \text{ и } e_j \text{ не является петлей,} \\ 2, & \text{если } v_i \text{ инцидентно } e_j, \text{ и } e_j \text{ – петля,} \\ 0 & \text{в остальных случаях.} \end{cases}$$

Замечание 2.2. Пусть $\Gamma = (V, E)$, $|\Gamma| = n$, простой сборный граф. Мы можем пронумеровать его вершины и ребра в порядке неубывания, в соответствии с трансверсальным путем от начальной до конечной вершины, начиная с 0. По лемме 1.17, имеем $V = \{0, 1, \ldots, n, n+1\}$ и $E = \{e_0, e_1, \ldots e_{2n-1}, e_{2n}\}$.

Определение 2.3. Матрицей инцидентности простого сборного графа Γ называется такая его матрица инцидентности $I(\Gamma)$, порядок вершин и ребер в которой зафиксирован в соответствии с замечанием 2.2.

Пример 2.4. На рисунке 5 изображен граф со сборным словом $w = 1122$, ребра которого пронумерованы в том порядке, в котором мы встречаем их, проходя по трансверсали.

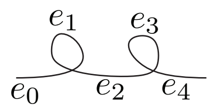


Рис. 5. Граф со сборным словом 1122.

Ниже представлена матрица инцидентности графа, изображенного на рисунке 5:

$$I(\Gamma_{1122}) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \!\!\begin{array}{c} \begin{array}{ccccc} e_0 & e_1 & e_2 & e_3 & e_4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Предложение 2.5. *Матрицы инцидентности простых сборных графов обладают следующими свойствами.*

(1) *Матрица инцидентности простого сборного графа с* $|\Gamma| = n$ *содержит* $n+2$ *строки и* $2n+1$ *столбец.*

(2) *Сумма элементов строки во всех строках, кроме первой и последней, равна* 4. *В первой и последней строках сумма элементов равна* 1.

(3) *Два столбца, не идущих подряд и имеющих общую строку с ненулевыми элементами, соответствуют двум ребрам, которые являются соседями в своей общей вершине.*

(4) *Два соседних столбца, соответствующих ребрам* e_k, e_{k+1}, *всегда имеют хотя бы одну общую строку с ненулевыми элементами.*

Доказательство. 1. Напрямую следует из леммы 1.17.

2. Справедливо, поскольку все вершины простого сборного графа, кроме начальной и конечной, имеют валентность 4.

3. Наличие общей строки с ненулевыми элементами эквивалентно тому, что у ребер есть общая вершина. Поскольку мы нумеровали ребра, двигаясь вдоль трансверсали, последовательные столбцы соответствуют ребрам, не являющимся соседями в общей вершине. Для вершины v и инцидентного ей ребра e существуют ровно два соседа ребра e в вершине v и только одно ребро, инцидентное v и не являющееся

соседом e в вершине v. Следовательно, два не идущих подряд столбца с общей ненулевой строкой соответствуют соседним ребрам.

3. Мы записываем ребра в том же порядке, в котором встречаем их, двигаясь вдоль трансверсали. Поэтому у двух ребер с подряд идущими номерами всегда есть общая вершина. □

Например, у столбцов матрицы $I(\Gamma_{1122})$, соответствующих ребрам e_1 и e_2, ненулевые элементы находятся только в строке, соответствующей вершине 1.

§3. Структурные графы и их матрицы

В этом параграфе мы рассмотрим серии графов, широко известные благодаря их экстремальным свойствам, см. [1,2,5]. Начнем с простых примеров сборных графов.

3.1. Примеры простых сборных графов. Начнем с графа, у которого нет 4-валентных вершин. Затем будем последовательно добавлять к нему петли, для того чтобы получить последовательность графов $\Gamma_\epsilon, \Gamma_{11}, \Gamma_{1122}, \Gamma_{112233}\ldots$, изображенную на рисунке 6.

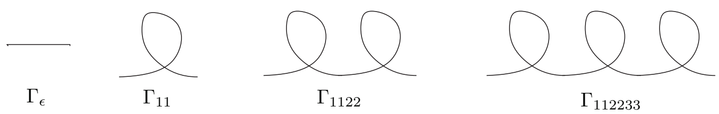


Рис. 6. Реализуемые сборные графы

Матрицы инцидентности этих графов имеют следующий вид:

$$I(\Gamma_\epsilon) = \begin{array}{c c} & e_0 \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 1 \\ 1 \end{pmatrix} \end{array}, \quad I(\Gamma_{11}) = \begin{array}{c c} & \begin{array}{ccc} e_0 & e_1 & e_2 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 1 \\ 0 & 0 & 1 \end{pmatrix} \end{array},$$

$$I(\Gamma_{1122}) = \begin{array}{c c} & \begin{array}{ccccc} e_0 & e_1 & e_2 & e_3 & e_4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array},$$

$$I(\Gamma_{112233}) = \begin{array}{cc} & \begin{array}{ccccccc} e_0 & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Легко видеть, что в каждой из строк, соответствующих вершинам v_k, $k \in \{1, \dots, n\}$, есть блок 121, начинающийся со столбца, соответствующего ребру $e_{2(k-1)}$.

3.2. Граф с заданным сборным числом. Покажем, как построить граф с произвольным сборным числом. Рассмотрим сборный граф со сборным словом $u = 122133$, см. рисунок 7.

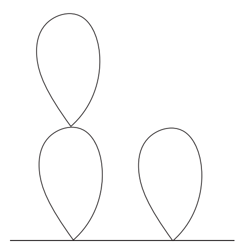


Рис. 7. Сборный граф Γ_u со сборным словом $u = 122133$.

Его матрица инцидентности имеет вид

$$I(\Gamma_u) = \begin{array}{cc} & \begin{array}{ccccccc} e_0 & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Далее, рассмотрим композицию $\Gamma_n := \Gamma_{aa} \circ \Gamma_u^n$. В [1] было установлено, что сборное число этой композиции $\mathrm{An}(\Gamma_n) = n + 1$. Таким образом, граф Γ_u помогает нам построить пример сборного графа с заданным сборным числом k. Заметим, что, по построению, $|\Gamma_k| = 3k + 1$.

На рисунках 8 и 9 изображены графы со сборными числами 2 и 3, построенные описанным выше способом. В §5 более детально описан процесс построения матрицы инцидентности для композиции двух графов.

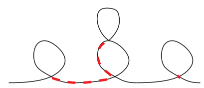


Рис. 8. Сборный граф $\Gamma_{aa} \circ \Gamma_u$ со сборным числом 2.

Матрица инцидентности этого графа имеет вид

$$I(\Gamma_{aa} \circ \Gamma_u) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \!\!\begin{array}{c} \begin{array}{ccccccccc} e_0 & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

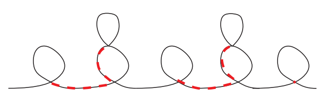


Рис. 9. Сборный граф $\Gamma_{aa} \circ \Gamma_u^2$ со сборным числом 3.

Матрица инцидентности этого графа имеет вид

$$I(\Gamma_{aa} \circ \Gamma_u^2) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{array} \!\!\begin{array}{c} \begin{array}{ccccccccccccccc} e_0 & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} & e_{13} & e_{14} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

3.3. Петельный граф.

Определение 3.1. Пусть n – натуральное число. Петельный граф размера n, обозначаемый TC_n, – это простой сборный граф, соответствующий слову, которое определяется индуктивно следующим образом: начальное слово есть $w_{TC_1} = 11$, далее каждое слово w_{TC_n} получается из $w_{TC_{n-1}}$ путем замены последней буквы $(n-1)$ на подслово $n(n-1)n$.

Для примера, первые четыре слова этой последовательности суть 11, 1212, 121323, 12132434, а n-тое слово это $w_{TC_n} = 121324354\ldots(n-1)(n-2)n(n-1)n$.

Рисунки 10, 11 и 12 ниже изображают серию примеров рассматриваемых графов при $n = 2, 3, 4$; помимо графов приведены соответствующие сборные слова и матрицы инцидентности.

Пример 3.2. Пусть $n = 2$. Тогда $w_{TC_2} = 1212$,

$$I(TC_2) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \!\!\begin{array}{c} \begin{array}{ccccc} e_0 & e_1 & e_2 & e_3 & e_4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array},$$

соответствующий граф изображен на рисунке 10.

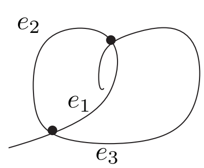


Рис. 10. TC_2 с помеченными ребрами.

Пример 3.3. Пусть $n = 3$. Тогда $w_{TC_3} = 121323$,

$$I(TC_3) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \!\!\begin{array}{c} \begin{array}{ccccccc} e_0 & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array},$$

соответствующий граф изображен на рисунке 11.

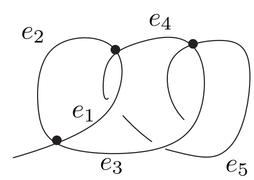


Рис. 11. TC_3 с помеченными ребрами.

Пример 3.4. Пусть $n = 4$. Тогда $w_{TC_4} = 12132434$,

$$I(TC_4) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccccccc} e_0 & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array},$$

соответствующий граф изображен на рисунке 12.

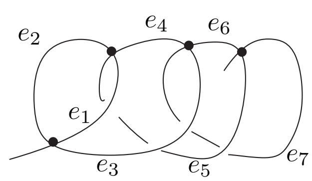


Рис. 12. TC_4 с помеченными ребрами.

На рисунке 13 представлен общий вид петельного графа и изображен процесс добавления веришины.

Предложение 3.5. *Для любого $n \geqslant 3$ строки матрицы $I(TC_n)$ имеют следующий вид:*

$$R_0(TC_n) = (1, \underbrace{0, \ldots, 0}_{2n \text{ раз}}), \quad R_1(TC_n) = (1, 1, 1, 1, \underbrace{0, \ldots, 0}_{2n-3 \text{ раз}}),$$

$$R_k(TC_n) = (\underbrace{0, \ldots, 0}_{2k-3 \text{ раз}}, 1, 1, 0, 1, 1, \underbrace{0, \ldots, 0}_{2n-2k-1 \text{ раз}}), \quad k = 2, \ldots, n-1,$$

$$R_n(TC_n) = (\underbrace{0,\ldots,0}_{2n-3\ \text{раз}},1,1,1,1), \quad R_{n+1}(TC_n) = (\underbrace{0,\ldots,0}_{2n\ \text{раз}},1).$$

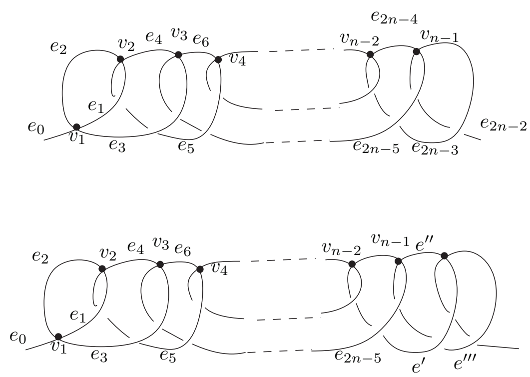


Рис. 13. Добавление вершины к TC_{n-1}.

Доказательство. Заметим, что строка, отвечающая вершине 0 (соответственно, $n+1$), содержит ровно один элемент 1 в первом (соответственно, последнем) столбце, а остальные элементы нулевые. Строка, соответствующая вершине 1 (соответственно, n) содержит единицы в четырех последовательных столбцах, начиная с e_0 (соответственно, с $e_{(2n-1)-2}$). Любой другой столбец, соответствующий вершине $k \in \{2,\ldots,n-1\}$, содержит блок $[1\,1\,0\,1\,1]$, начинающийся со столбца, отвечающего ребру e_{2k-3}. Простой подсчет элементов завершает доказательство. $\square$

3.4. Возвратные слова.

Определение 3.6. Возвратное слово – это 2-слово, имеющее вид

$$a_1a_2\ldots a_{n-1}a_na_na_{n-1}\ldots a_2a_1.$$

Далее возвратное слово с n различными буквами мы будем обозначать $\mathrm{ret}(n)$.

Возвратные слова используются для представления частей генома микронуклеуса, соответствующих часто встречающимся последовательностям, и играют важную роль в изучении индекса вложенности, более подробно см. [4] и [3].

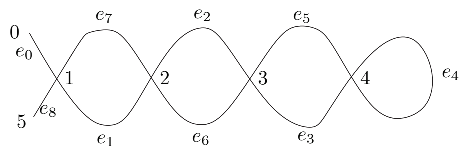


Рис. 14. Сборный граф, соответствующий слову ret(4).

Пример 3.7. Рассмотрим, например, 2-слово 12344321. Соответствующий ему сборный граф изображен на рисунке 14.

Матрица инцидентности этого графа имеет вид

$$I(\Gamma_{ret(4)}) = \begin{array}{c} \\ 0\\1\\2\\3\\4\\5 \end{array} \begin{array}{c} \begin{array}{ccccccccc} e_0 & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Предложение 3.8. *Для любого $n \geqslant 2$ строки матрицы $I(\Gamma_{ret(n)})$ имеют следующий вид:*

$$R_0(\mathrm{ret}(n)) = (1, \underbrace{0, \ldots, 0}_{2n \text{ раз}})$$

$$R_k(\mathrm{ret}(n)) = (\underbrace{0, \ldots, 0}_{k-1 \text{ раз}}, 1, 1 \underbrace{0, \ldots, 0}_{2n-2k-1 \text{ раз}}, 1, 1, \underbrace{0, \ldots, 0}_{k-1 \text{ раз}}), \quad k = 1, \ldots, n-1,$$

$$R_n(\mathrm{ret}(n)) = (\underbrace{0, \ldots, 0}_{n-1 \text{ раз}}, 1, 2, 1, \underbrace{0, \ldots, 0}_{n-1 \text{ раз}}),$$

$$R_{n+1}(\mathrm{ret}(n)) = (\underbrace{0, \ldots, 0}_{2n \text{ раз}}, 1).$$

Иными словами, если исключить первую и последнюю строки, то матрица инцидентности графа со сборным словом ret(n) *симметрична относительно столбца, индексированного ребром e_n.*

Доказательство. Первая и последняя строки соответствуют начальной и конечной вершинам, каждой из которых инцидентно ровно одно ребро, а, значит, в каждой из строк имеется ровно один ненулевой элемент. Единственная вершина, в которой есть петля, – это вершина с номером n, а петля находится в середине эйлеровой трансверсали, т.е. это ребро e_n. Очевидно, ребра e_{n-1} и e_{n+1} также инцидентны вершине n. Поскольку возвратное слово – палиндром, если пройти по эйлеровой трансверсали в обратном направлении ничего не изменится. Следовательно, e_k и e_{2n-k}, соединяют одни и те же вершины. Очевидно, что $e_k,\ k = 0, 1, \ldots, n-1$, соединяет вершины с номерами k и $k+1$. □

3.5. Повторные слова.

Определение 3.9. Повторное слово – это 2-слово, имеющее вид

$$a_1a_2 \ldots a_{n-1}a_na_1a_2 \ldots a_{n-1}a_n.$$

Заметим, что любое повторное слово из n различных букв эквивалентно $123 \ldots n123 \ldots n$. Поэтому далее мы обозначаем повторное слово с n различными буквами rep(n).

Так же как и возвратные слова, повторные слова помогают исследовать сложность генов микронуклеуса, более подробно см. в [4] и [3].

Пример 3.10. Рассмотрим слово rep(4) = 12341234. Соответствующий граф изображен на рисунке 15.

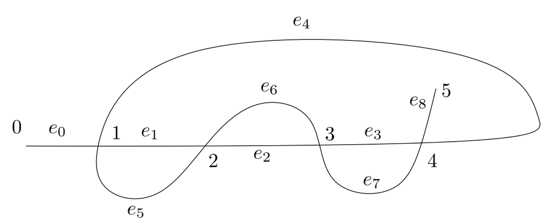


Рис. 15. Сборный граф, соответствующий слову rep(4).

Матрица инцидентности этого графа есть

$$I(\Gamma_{rep(4)}) = \begin{array}{c c} & \begin{array}{ccccccccc} e_0 & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Предложение 3.11. *Для любого $n \geqslant 2$ строки матрицы $I(\Gamma_{\mathrm{rep(n)}})$ имеют следующий вид:*

$$R_0(\mathrm{rep}(n)) = (1, \underbrace{0, \dots, 0}_{2n \text{ раз}})$$

$$R_k(\mathrm{rep}(n)) = (\underbrace{0, \dots, 0}_{k-1 \text{ раз}}, 1, 1\, \underbrace{0, \dots, 0}_{n-2 \text{ раз}}, 1, 1, \underbrace{0, \dots, 0}_{n-k \text{ раз}}), \quad k = 1, \dots, n$$

$$R_{n+1}(\mathrm{rep}(n)) = (\underbrace{0, \dots, 0}_{2n \text{ раз}}, 1).$$

Доказательство. Рассмотрим цепь ребер $e_0, \dots, e_{n-1}$, соответствующих первым n столбцам нашей матрицы. Ребро e_i, $i \in \{0, 1, \dots, n-1\}$, соединяет вершины i и $i+1$. Отсюда следует, что элементы $R_i(\mathrm{rep}(n))[i]$ (i-тый элемент i-той строки) и $R_{i+1}(\mathrm{rep}(n))[i]$ (i-тый элементы $(i+1)$-ой строки) равны 1 для $i \in \{0, 1, \dots, n-1\}$. Ребро e_n соединяет вершины n и 1, откуда $R_1(\mathrm{rep}(n))[n] = 1$ и $R_n(\mathrm{rep}(n))[n] = 1$. Далее, рассмотрим ребра $e_{n+1}, \dots, e_{2n}$, соответствующие последним n столбцам. Ребро $e_i, i \in \{n+1, n+2, \dots, 2n\}$, соединяет вершины $i-n$ и $i-n+1$. Поэтому элементы $R_i(\mathrm{rep}(n))[n+i]$ и $R_{i+1}(\mathrm{rep}(n))[n+i]$ равны 1 при $i \in \{n+1, n+2, \dots, 2n\}$. Итак, матрица инцидентности нашего графа имеет описанный выше вид. □

§4. Внутренняя петельная подстановка

Внутренняя петельная подстановка – это важное преобразование сборного графа, см. [5] для более подробной информации. В этом параграфе мы описываем соответствующие преобразования матриц. Начнем с необходимых определений.

Определение 4.1. *Петля* это сборный граф со сборным словом 11, см. рисунок 16.

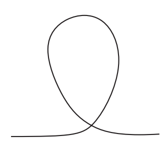


Рис. 16. Петля $\Gamma_{(1)}$.

Лемма 4.2 ([5, стр. 15, определение 5.1])**.** *Пусть* Γ *– сборный граф, а* Γ' *получается из* Γ *заменой одного из ребер на петлю* $\Gamma_{(1)}$*. Тогда* Γ' *– сборный граф.*

Доказательство. Легко видеть, что все вершины Γ' имеют валентность 1 или 4, и число вершин валентности 1 осталось прежним. Покажем теперь, что в Γ' есть эйлеров трансверсальный путь. Рассмотрим эйлеров трансверсальный путь в Γ. Ребро, которое заменили на петлю, – часть этой трансверсали. Тогда можно включить рассматриваемую петлю в уже существующую эйлерову трансверсаль и получить нужный путь в Γ'. □

Рисунок 17 иллюстрирует включение петли в трансверсаль в двух стандартных случаях, когда петля добавляется на другую петлю и на ребро, петлей не являющееся.

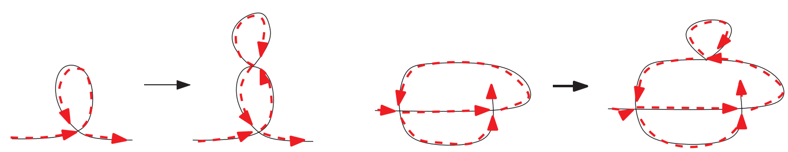


Рис. 17. Включение петли в имеющуюся трансверсаль.

Применив последнюю лемму несколько раз, можно получить, что замена нескольких ребер графа на петли дает сборный граф. Последнее обстоятельство приводит нас к следующему определению.

Определение 4.3 ([5, определение 5.1])**.** Пусть Γ – сборный граф, а $\tilde{\Gamma}$ получен из Γ заменой каждого ребра на петлю $\Gamma_{(1)}$. Тогда граф $\tilde{\Gamma}$ называют *графом петельной подстановки* или графом, полученным петельной подстановкой. Сборный граф $\tilde{\Gamma}^{\circ}$ называется *графом внутренней петельной подстановки*, если он получается из Γ подстановкой петли $\Gamma_{(1)}$ вместо каждого из ребер, кроме двух ребер, инцидентных концевым вершинам.

Матрицу инцидентности графа (внутренней) петельной подстановки легко построить по матрице инцидентности исходного графа.

Предложение 4.4. *Пусть* Γ *– сборный граф и пусть* $\Gamma^{(i)}$ *получен из* Γ *заменой* i*-того ребра на петлю* $\Gamma_{(1)}$*. Тогда верны следующие утверждения.*

1. $I(\Gamma^{(i)})$ получается из $I(\Gamma)$, если вместо i-того столбца подставить три столбца C_1, C_2, C_3 и добавить строку R описанным ниже способом.

Пусть ненулевые элементы столбца c_i расположены в k-той и l-той строках и $k \leqslant l$. При $k = l$ строка R вставляется под строкой с номером k. При $k \neq l$ строка R вставляется над строкой с номером l.

Элементы вставляемых столбцов таковы: в строке R столбца C_2 стоит элемент 2, в столбце C_1 единицы стоят в строках R и k, столбец C_3 содержит 1 в строке R и 1 в строке, которая ранее была l-той. Все остальные элементы добавленных строк и столбцов равны 0.

2. $I(\Gamma)$ получается из $I(\Gamma^{(i)})$ вычеркиванием строки и столбца, соответствующих $\Gamma_{(1)}$, и заменой двух столбцов, у которых в вычеркнутой строке стоят единицы, на их сумму.

Доказательство. 1. Добавление петли в граф добавляет два ребра и вершину. По определению, петля соответствует столбцу, в котором в добавленной строке стоит элемент 2. Ребро между вершинами k и l оказывается разделено добавленной вершиной на два ребра, а потому, соответствующий столбец должен быть заменен на два столбца: с единицами на k-том месте и в новой строке (C_1), и единицами на l-том месте и в новой строке (C_3). Также нужно добавить между ними новый столбец C_2, который соответствует петле.

2. Наоборот, если $\Gamma_{(1)}$ – петля, стоящая на i-том ребре, то строка, соответствующая вершине $\Gamma_{(1)}$, содержит ровно три ненулевых элемента: 1, 1, 2. Необходимо удалить столбец, содержаший элемент 2, а два оставшихся заменить либо на столбец с единицами в тех строках, в которых они стояли, если это разные строки, либо на столбец с элементом 2, если единицы в заменяемых столбцах стояли в одной и той же строке. Остальные элементы остаются нулевыми. Очевидно, что подставляемый столбец – это как раз сумма удаляемых столбцов. □

Следствие 4.5. *Пусть Γ – сборный граф. Тогда матрицу инцидентности $I(\tilde{\Gamma})$ (соответственно, $I(\tilde{\Gamma}^{\circ})$) графа петельной (соответственно, внутренней петельной) подстановки можно получить из $I(\Gamma)$, применяя процедуру из предложения 4.4 последовательно к каждому столбцу (соответственно, к каждому столбцу кроме первого и* *последнего). Аналогичным образом можно проделать обратную процедуру. Заметим, что если* $|\Gamma| = n$, *то* $I(\Gamma) \in M_{n+2,2n+1}$, $I(\tilde{\Gamma}) \in M_{3n+3,6n+3}$, $I(\tilde{\Gamma}^{\circ}) \in M_{3n+1,6n-1}$.

Пример 4.6. Пусть Γ – граф, изображенный на рисунке 18.

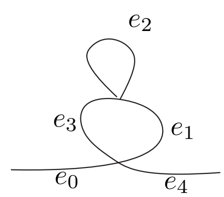


Рис. 18. $\Gamma = \Gamma_{1221}$ с помеченными ребрами.

Тогда Γ – сборный граф со сборным словом $w = 1221$. Соответствующий граф петельной подстановки $\tilde{\Gamma}$ изображен на рисунке 19. Матрицы инцидентности этих графов имеют вид:

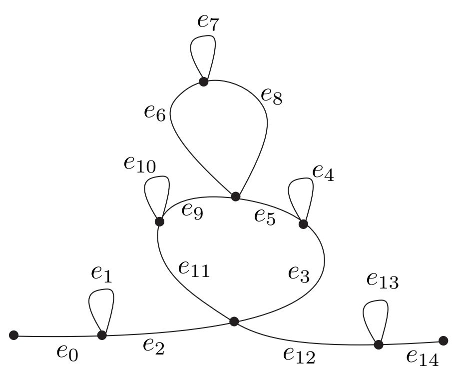


Рис. 19. Граф петельной подстановки $\tilde{\Gamma}$.

$$I(\tilde{\Gamma}) = \begin{array}{c} \\ 0\\1\\2\\3\\4\\5\\6\\7\\8 \end{array} \begin{array}{c} \begin{array}{ccccccccccccccc} e_0 & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} & e_{13} & e_{14} \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

и

$$I(\Gamma_{1221}) = \begin{array}{c c} & \begin{array}{ccccc} e_0 & e_1 & e_2 & e_3 & e_4 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Ниже мы приведем алгоритм, который позволяет построить матрицу исходного сборного графа по матрице графа петельной подстановки. Для упрощения записей будем считать, что число вершин в графе петельной подстановки равно $3n + 3$, а число вершин в графе внутренней петельной подстановки равно $3n + 1$. В таком случае, у графа петельной подстановки $6n + 3$ ребер, а у графа внутренней петельной подстановки, соответственно, $6n - 1$ ребро.

Алгоритм 4.7.

1. *Пометим все строки, которые соответствуют петлям.*
2. *Заменим тройку столбцов с номерами $k, k + 1, k + 2$ на столбец, представляющий собой векторную сумму столбцов с индексами $k, k + 2$. Здесь $k = 0, 3, 6, \ldots, 6n$ в случае петельной подстановки и $k = 1, 4, 7, \ldots, 6n - 5$ в случае внутренней петельной подстановки.*
3. *Удалим строки, отмеченные на шаге* 1.

Доказательство. Данный алгоритм основан на том, что каждое вхождение числа 2 в матрицу инцидентности относится к петле, добавленной при петельной подстановке. Таким образом, мы должны удалить все строки и столбцы, содержащие элемент 2, и склеить столбцы $k, k + 2$, относящиеся к одному и тому же ребру. Мы одинаково действуем для каждой тройки столбцов, начиная с 0-го столбца в случае петельной подстановки и с 1-го столбца в случае внутренней петельной подстановки. □

Пример 4.8. Чтобы проиллюстрировать алгоритм, приведенный выше, рассмотрим граф петельной подстановки, представленный на рисунке 19. Для начала пометим все строки, соответствующие петлям:

$$
I(\Gamma) = \begin{array}{c} \\ 0 \\ \mathbf{1} \\ 2 \\ \mathbf{3} \\ 4 \\ \mathbf{5} \\ \mathbf{6} \\ \mathbf{7} \\ 8 \end{array}
\begin{array}{c}
\begin{array}{ccccccccccccccc} e_0 & e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 & e_8 & e_9 & e_{10} & e_{11} & e_{12} & e_{13} & e_{14} \end{array} \\
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 2 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}.
$$

Затем выполним второй шаг и заменим столбцы на сумму векторов:

$$
\begin{array}{c} \\ 0 \\ \mathbf{1} \\ 2 \\ \mathbf{3} \\ 4 \\ \mathbf{5} \\ \mathbf{6} \\ \mathbf{7} \\ 8 \end{array}
\begin{array}{c}
\begin{array}{ccccc} e_0+e_2 & e_3+e_5 & e_6+e_8 & e_9+e_{11} & e_{12}+e_{14} \end{array} \\
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
2 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 \\
0 & 2 & 0 & 0 & 0 \\
0 & 1 & 2 & 1 & 0 \\
0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 2 & 0 \\
0 & 0 & 0 & 0 & 2 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}.
$$

Наконец, удалим отмеченные строки и получим матрицу инцидентности для графа со сборным словом $w = 2442$:

$$
I(\Gamma') = \begin{array}{c} \\ 0 \\ 2 \\ 4 \\ 8 \end{array}
\begin{array}{c}
\begin{array}{ccccc} e_0+e_2 & e_3+e_5 & e_6+e_8 & e_9+e_{11} & e_{12}+e_{14} \end{array} \\
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 1 \\
0 & 1 & 2 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}.
$$

Очевидно, что матрица инцидентности $I(\Gamma_{1221})$ графа, представленного на рисунке 18, и матрица $I(\Gamma')$ совпадают.

§5. Матрица инцидентности для композиции

Матрица инцидентности композиции получается как блочная матрица с блоками, соответствующими матрицам инцидентности компонент, за исключением строк, относящихся к конечной (начальной) вершинам.

Предложение 5.1. *Пусть Γ_1 и Γ_2 – два простых сборных графа с матрицами инцидентности $I(\Gamma_1)$ и $I(\Gamma_2)$ порядков $|\Gamma_1|=n$ и $|\Gamma_2|=m$*

соответственно. Тогда матрица инцидентности композиции $I(\Gamma_1 \circ \Gamma_2)$*, содержащая* $n+m+2$ *строки и* $2n+2m+1$ *столбец, имеет вид (считаем, что строки и столбцы индексируются с* 0*):*

1. $I(\Gamma_1 \circ \Gamma_2)_{k,l} = I(\Gamma_1)_{k,l}$, *если* $0 \leqslant k \leqslant n$, $0 \leqslant l \leqslant 2n$,
2. $I(\Gamma_1 \circ \Gamma_2)_{k,l} = 0$, *если* $n < k \leqslant m+n+1$ *и* $0 \leqslant l < 2n$ *или если* $0 \leqslant k < n+1$ *и* $2n < l \leqslant 2m+2n$,
3. $I(\Gamma_1 \circ \Gamma_2)_{k,l} = I(\Gamma_2)_{k-n,l-2n}$, *если* $n+1 \leqslant k \leqslant m+n+1, 2n \leqslant l \leqslant 2n+2m$.

Доказательство. Действительно, вершины композиции графов делятся на две части: вершины Γ_1, исключая конечную, и вершины Γ_2, исключая начальную. Внутри каждой из частей вершины связаны ребрами так же, как и в исходных графах. Ребер между вершинами из разных частей нет, за исключением ребра e_{2n}, соответствующего столбцу $2n$, которое соединяет две части. □

Чтобы проиллюстрировать это предложение, рассмотрим, например, два графа со сборными словами 1212 и 1221, см. рис. 20.

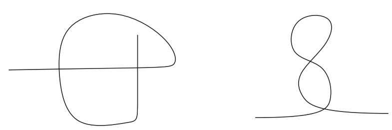


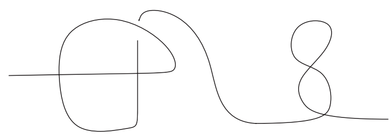
Рис. 20. Графы до композиции.

Матрицы инцидентности этих графов таковы:

$$I(TC_2) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \!\!\begin{array}{c} \begin{array}{ccccc} e_0 & e_1 & e_2 & e_3 & e_4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}, \quad I(\Gamma_{1221}) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \!\!\begin{array}{c} \begin{array}{ccccc} e_0 & e_1 & e_2 & e_3 & e_4 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$

Матрица инцидентности композиции $1212 \circ 1221$ строится следующим способом. Возьмем $I(TC_2)$ и удалим последнюю строку. Затем возьмем $I(\Gamma_{1221})$ и удалим первую строку. После этого запишем $I(\Gamma_{1221})$ внизу, начиная со столбца, который соответствует ребру, инцидентному конечной вершине первого графа, а остальные элементы заполним нулями:

	e_0	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
0	1	0	0	0	0	0	0	0	0
1	1	1	1	1	0	0	0	0	0
2	0	1	1	1	1	0	0	0	0
. 3	0	0	0	0	1	1	0	1	1
4	0	0	0	0	0	1	2	1	0
5	0	0	0	0	0	0	0	0	1



1212 1221

Рис. 21. Графы после композиции.

Список литературы

1. A. Angeleska, N. Jonoska, M. Saito, *DNA recombinations through assembly graphs.* — Discr. Appl. Math. **157** (2009), 3020–3037.
2. A. Angeleska, N. Jonoska, M. Saito, L. F. Landweber, *RNA-guided DNA assembly.* — J. Theor. Biology **248** (4) (2007), 706–720.
3. R. Arredondo, *Properties of graphs used to model DNA recombination.* — Grad. Theses and Dissertations (2014). `https://scholarcommons.usf.edu/etd/4979`.
4. R. Arredondo, *Reductions on double occurrence words.* — `arXiv:1311.3543`, November, 2013.
5. J. Burns, E. Dolzhenko, N. Jonoska, T. Muche, M. Saito, *Four-regular graphs with rigid vertices associated to DNA recombination.* — Discr. Appl. Math. **161** (2013), 1378–1394.
6. A. Ehrenfeucht, T. Harju, I. Petre, D. M. Prescott, G. Rozenberg, *Computing in Living Cells.* Springer, 2005.
7. A. Ehrenfeucht, T. Harju, G. Rozenberg, *Gene assembly through cyclic graph decomposition.* — Theor. Comput. Sci. **281** (2002), 325–349.
8. L. Kari, L. F. Landweber, *Computational power of gene rearrangement.* in: DNA Based Computers (E. Winfree, D.K. Gifford eds.), AMS (1999), pp. 207–216.
9. M. Nowacki, V. Vijayan, Y. Zhou, K. Schotanus, T. G. Doak, L. F. Landweber, *RNA-mediated epigenetic programming of a genome-rearrangement pathway.* — Nature **451** (2008), 153–159.
10. D. M. Prescott, A. Ehrenfeucht, G. Rozenberg, *Template-guided recombination for IES elimination and unscrambling of genes in stichotrichous ciliates.* — J. Theor. Biology **222** (2003), 323–330.

Guterman A. E., Kreines E. M., Ostroukhova N. V. 2-words: their graphs and matrices.

Double-occurrence words play an important role in genetics for describing epigenetic genome rearrangements. A useful geometric representation for double-occurrence words is given by the so-called assembly graphs. The paper investigates properties of the incidence matrices that correspond to assembly graphs. An explicit matrix characterization for simple assembly graphs of a given structure and series of constructions, using these graphs and important for genetic investigations, are provided.

Поступило 7 октября 2019 г.

Московский государственный
университет им. М. В. Ломоносова,
Москва 119991, Россия;
Московский физико-технический институт,
Долгопрудный 141701, Россия

E-mail: guterman@list.ru

Московский государственный
университет им. М. В. Ломоносова
Москва 119991, Россия;
Московский физико-технический институт,
Долгопрудный, 141701, Россия

E-mail: elena.kreines@gmail.com

Московский государственный
университет им. М. В. Ломоносова
Москва 119991, Россия

E-mail: natosova@gmail.com