Содержание

1 Аннотация

Проект посвящен 2-словам, которые играют важную роль в генетике при описании эпигенетических геномных перестроек. Удобным геометрическим представлением 2-слов являются так называемые сборные графы. Среди их характеристик выделяется сборное число – минимальное количество полигональных путей, покрывающих все 4-валентные вершины графа. В данном проекте будет сделан упор на 2-слова, являющиеся палиндромами и полупалиндромами: будут исследоваться их комбинаторные свойства и характеристики. Один из интересующих вопросов таков: верно ли, что сборное число любого полупалиндрома равно 1?

2 Обзор источников

Статья[1] содержит определение таких понятий, как сборный граф, изоморфизм сборных графов, трансверсаль, простой сборный граф, слово, сборное слово или 2-слово, разложимый и неразложимый сборный граф, Гамильтоново множество путей, полигональный путь, сборное число, реализуемый и нереализуемый сборный граф, минимальное реализующее число, а также некоторые теоремы.

Последующая статья по теме[2] определяет такие понятия, как 2-слово в порядке возрастания, палиндром, сильно-неразложимый сборный граф, а также выводит формулы для подсчета всех, неразложимых и сильно-неразложимых 2-слов, а также всех, неразложимых и сильно-неразложимых палиндромов.

В статье[3] рассматриваются матрицы инцедентности простых сборных графов и их свойства, серии сборных графов, широко известные благодаря их экстремальным свойствам, а также внутренняя петельная подстановка.

Рассматриваются конечные графы $\Gamma = (V, E)$, где V — множество вершин, а $E \subseteq V \times V$ — множество ребер. Граф может содержать петли и кратные ребра.

Определение 1. Степень вершины $v \in V$ — число ребер, инцидентных данной вершине.

Определение 2. Циклический порядок для кортежа из k элементов $(x_1, x_2, x_3, \ldots, x_{k-1}, x_k)$ — множество

$$\begin{aligned}&(x_1, x_2, x_3, \ldots, x_{k-1}, x_k)^{cyc} = \{(x_1, x_2, x_3, \ldots, x_{k-1}, x_k),\\ &(x_2, x_3, \ldots, x_{k-1}, x_k, x_1), (x_3, \ldots, x_{k-1}, x_k, x_1, x_2), \ldots,\\ &(x_k, x_1, x_2, x_3, \ldots, x_{k-1}), (x_k, x_{k-1}, x_{k-2}, \ldots, x_2, x_1),\\ &(x_{k-1}, x_{k-2}, \ldots, x_2, x_1, x_k), (x_{k-2}, \ldots, x_2, x_1, x_k, x_{k-1}), \ldots,\\ &(x_1, x_k, x_{k-1}, x_{k-2}, \ldots, x_2)\}\end{aligned}$$

то есть все циклические сдвиги кортежа и все циклические сдвиги кортежа, записанного в обратном порядке.

Чтобы задать циклический порядок, достаточно одного элемента множества $(x_1, x_2, x_3, \ldots, x_{k-1}, x_k)^{cyc}$.

Определение 3. Вершина v — упорядоченная (или, иногда, регулярная), если циклический порядок ребер, инцидентных ей, зафиксирован.

Замечание 4. Для каждого из ребер упорядоченной вершины корректно определены его соседи.

Определение 5 ([1, Definition 3.1.])**.** Сборный граф — конечный связный граф, в котором все вершины упорядоченные и имеют степени 1 или 4.

Определение 6. Концевая вершина — вершина степени 1.

Определение 7. Порядок Γ (обозначение $|\Gamma|$) — количество вершин, степени 4 сборного графа Γ.

Определение 8. Тривиальный сборный граф — сборный граф Γ, что $|\Gamma| = 0$.

Определение 9 ([1, Definition 3.2.])**.** Графы Γ_1 и Γ_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 $\phi(u)$.

Определение 10. Последовательность вершин и ребер $(v_0, e_1, v_1, e_2, \dots, e_n, v_n)$ — путь в графе, если $v_i \in V$, $e_i \in E$, e_i — ребро между v_{i-1} и v_i.

Определение 11 ([1, Definition 3.3.])**.** Трансверсаль $(v_0, e_1, v_1, e_2, \dots, e_n, v_n)$ — путь, что каждая вершина встречается максимум два раза, все ребра различны, а ребра e_i и e_{i+1} не являются соседними в v_i.

Определение 12. Эйлеров путь в Γ — путь, содержащий каждое ребро ровно один раз.

Определение 13 ([1, Definition 3.5.])**.** Простой сборный граф — сборный граф, содержащий эйлерову трансверсаль.

Определение 14. Две трансверсали эквивалентны, или если они равны, или если одна является другой в обратном порядке.

Лемма 15 ([1, Lemma 3.6.])**.** *В простом сборном графе существует единственный класс эквивалентности трансверсалей.*

Лемма 16 ([1, Lemma 3.7.])**.** *Два простых сборных графа Γ_1 и Γ_2 с трансверсалями γ_1 и γ_2 изоморфны если и только если существует отображение $\Phi = (\Phi_v, \Phi_e) : \Gamma_1 \to \Gamma_2$ с биекциями $\Phi_v : V_1 \to V_2$ и $\Phi_e : E_1 \to E_2$, что $\Phi(\gamma_1)$ эквивалентна γ_2.*

Сборные графы естественным образом связаны со специальным классом слов.

Определение 17. Сборное слово или 2-слово — это слово в некотором алфавите $S = \{a_1, a_2, \dots\}$, что каждая буква a_i либо содержится в слове ровно два раза, либо не содержится вовсе.

Определение 18. Порядок 2-слова w (обозначение $|w|$) — количество букв, которые встречаются в 2-слове.

Определение 19. Обратное слово к слову $w = a_{i_1} \dots a_{i_k}$ (обозначение w^R) — $a_{i_k} \dots a_{i_1}$.

Определение 20. Два 2-слова эквивалентны, если после переименования некоторых букв они или совпадают, либо являются обратными друг для друга.

Лемма 21 ([1, Lemma 3.8.])**.** *Классы эквивалентности 2-слов находятся в биективном соответствии с классами изоморфизма простых сборных графов.*

Определение 22 ([1, Definition 3.10.])**.** 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 23. Композиция простых сборных графов — простой сборный граф.

Определение 24 ([1, Definition 3.11.])**.** Разложимое 2-слово w — такое 2-слово, которое может быть записано как произведение $w = uv$ двух непустых 2-слов u, v. Аналогично, разложимый простой сборный граф Γ — такой сборный граф, что $\Gamma = \Gamma_1 \circ \Gamma_2$ для непустых простых сборных графов Γ_1, Γ_2. В противном случае и 2-слово, и простой сборный граф — неразложимые.

Далее рассматриваются полигональные пути и сборное число.

Определение 25. Простой путь — путь, не содержащий какую-либо вершину дважды.

Определение 26 ([1, Definition 4.2.])**.** Множество попарно непересекающихся простых путей $\{\gamma_1, \dots, \gamma_k\}$ — гамильтоново, если их объединение содержит все вершины степени 4 графа Γ.

Определение 27 ([1, Definition 4.3.]). Полигональный путь — путь $(v_0, e_1, v_1, e_2, \ldots, e_n, v_n)$, что e_i e_{i+1} — соседи для v_i для $i \in \{1, \ldots, n-1\}$.

Определение 28 ([1, Definition 4.4.]). Сборное число простого сборного графа Γ (обозначение $\mathrm{An}(\Gamma)$), определяется как $\mathrm{An}(\Gamma) = \min\{k|$ существует гамильтоново множество полигональных путей $\{\gamma_1, \ldots, \gamma_k\}$ в $\Gamma\}$.

Определение 29. Реализумый простой сборный граф — простой сборный граф, со сборным числом 1. Иначе — нереализуемый.

Лемма 30 ([1, Lemma 4.6.]). *Для любой пары ориентированных простых сборных графов* Γ_1 *и* Γ_2*, одно из двух равенств выполнено:* $\mathrm{An}(\Gamma_1 \circ \Gamma_2) = \mathrm{An}(\Gamma_1) + \mathrm{An}(\Gamma_2)$*, или* $\mathrm{An}(\Gamma_1 \circ \Gamma_2) = \mathrm{An}(\Gamma_1) + \mathrm{An}(\Gamma_2) - 1$.

Предложение 31 ([1, Proposition 4.9.]). *Для любого натурального числа* n

1. *Существует разложимый сборный граф* Γ*, что* $\mathrm{An}(\Gamma) = n$.
2. *Существует неразложимый сборный граф* Γ*, что* $\mathrm{An}(\Gamma) = n$.

Определение 32 ([1, Definition 4.10.]). Минимальное реализующее число для натурального числа n (обозначение $R_{\min}(n)$) определяется как $R_{\min}(n) = \min\{|\Gamma| : \mathrm{An}(\Gamma) = n\}$. Граф Γ, такой что $R_{\min}(n) = |\Gamma|$, — реализация $R_{\min}(n)$.

Предложение 33 ([1, Proposition 4.13.]). *Следующие свойства выполняются для* $R_{\min}$

1. *Для любого натурального числа* n $R_{\min}(n) < R_{\min}(n+1)$.
2. *Если* $R_{\min}(n) = k$*, то* $\forall s \geq k$ *существует сборный граф* Γ*, что* $|\Gamma| = s, \mathrm{An}(\Gamma) = n$.
3. *Для любого натурального числа* n $R_{\min}(n) \leq 3(n-1) + 1$.

Предложение 34 ([1, Proposition 4.14.]). *Существует константа* N *такая, что для любого реализуемого 2-слова* w*, существует нереализуемое неразложимое* v*, что* $w \subset v$ *и* $|v| - |w| < N$.

Предложение 35 ([1, Proposition 4.17.]). *Для любого нереализуемого 2-слова* v*, что* $|v| = m$*, существует константа* $N(m)$ *и реализуемое 2-слово* w*, что* $v \subset w$ *и* $|w| - |v| \leq N(m)$.

Определение 36. Пусть слова записаны над алфавитом линейно сравнимых элементов. Тогда говорят, что слово записано в порядке возрастания, если i-ая буква в слове по величине встречается в слове первый раз только после всех букв, меньших ее.

Далее мы отождествляем 2-слово и его запись в возрастающем порядке.
Далее рассматриваются комбинаторные свойства 2-слов.

Лемма 37 ([2, Lemma 3.2.]). *Мощность множества 2-слов на* n *буквах есть*

$$W_n = (2n-1)!!$$

Определение 38 ([2, Definition 3.3.]). Палиндром — такое 2-слово, что его обратное (записанное в возрастающем порядке) равно ему.

Лемма 39 ([2, Lemma 3.4.]). *Количество палиндромов на* n *буквах есть*

$$P_n = \sum_{k=\lfloor n/2 \rfloor}^{n} \binom{k}{n-k} \frac{n!}{k!}$$

Лемма 40 ([2, Lemma 3.6.]). *Количество неразложимых 2-слов на n буквах есть*

$$I_1 = 1; \qquad I_n = W_n - \sum_{k=1}^{n-1} W_k I_{n-k}$$

Лемма 41 ([2, Lemma 3.7.]). *Количество неразложимых палиндромов на n буквах есть*

$$J_1 = 1; \qquad J_n = P_n - \sum_{k=1}^{\lfloor n/2 \rfloor} W_k J_{n-2k}$$

Определение 42 ([2, Definition 3.8.]). Сильно-неразложимое 2-слово — такое 2-слово, что оно не содержит никакого собственного 2-подслова.

Лемма 43 ([2, Proposition 3.10]). *Количество сильно-неразложимых 2-слов на n буквах есть*

$$S_1 = 1; \qquad S_n = (n-1) \sum_{i=1}^{n-1} S_i S_{n-i}$$

Лемма 44 ([2, Proposition 3.10]). *Количество сильно-неразложимых палиндромов на n буквах есть*

$$T_0 = -1; \qquad T_1 = 1; \qquad T_n = (n-1) \sum_{i=1}^{n-2} T_i T_{n-i} + \sum_{i=1}^{\lfloor n/2 \rfloor} (2n - 4i - 1) S_i T_{n-2i}$$

3 Полученные результаты

3.1 Теоретические

Определение 45. 2-слово w на n буквах в возрастающем порядке — полупалиндром, если $\forall i \in \{1, \dots, 2n\} : w_{2n-i+1} = n - w_i + 1$.

Пример 46. 1122 и 1212 — полупалиндромы, а 1221 — нет.

Определение 47. Скобочная последовательность — символьная последовательность, состоящая из символов “(” и “)”

Определение 48 ([4, 2.10.1 Правильные последовательности скобок])**.** Правильная скобочная последовательность (ПСП) определяется индуктивно

- пустая последовательность правильна;
- если A — правильная скобочная последовательность, то (A) — правильная скобочная последовательность;
- если A, B — правильные скобочные последовательности, то их конкатенация AB — правильная скобочная последовательность.

Определение 49. Префикс ПСП длины n — скобочная последовательность, которую можно дополнить до правильной.

Определение 50. Баланс скобочной последовательности — разность количества открывающихся скобок и количества закрывающихся.

Лемма 51. *Аналогично [4, Теорема 2.2.] скобочная последовательность является префиксом ПСП тогда и только тогда, когда на любом ее префиксе баланс неотрицателен.*

Определение 52. Симметричная ПСП — такая ПСП, что если на i месте стоит открывающаяся скобка, то на $n - i + 1$ стоит закрывающаяся, а если закрывающаяся, то открывающаяся.

Замечание 53. Можно провести естественное отображение между 2-словами в возрастающем порядке и ПСП: первому вхождению символа сопоставить открывающуюся скобку, второму вхождению — закрывающуюся. Это отображение не является инъекцией: 1212 и 1221 переходят в $(())$. При этом это отображение — сюръекция.

Предложение 54. *Множество полупалиндромов на n буквах биективно множеству симметричных ПСП длины $2n$.*

Доказательство. Рассмотрим отображение f из множества симметричных ПСП в множество полупалиндромов: отдельно пронумеруем открывающиеся скобки в порядке возрастания, отдельно закрывающиеся в порядке возрастания и запишем это в строку. Примеры: $()()() \to 112233$, $(()()) \to 121323$,$((())) \to 123123$. Заметим, что скобки, которые при анализе ПСП разбиваются на пары “открывающая-закрывающая” не соответствуют парам букв.

Покажем, что оно действительно бьет в множество полупалиндромов. Для начала, результат отображения — 2-слово, так как количество открывающихся и закрывающихся скобочек одинаково. Результат в возрастающем порядке — первый раз каждое число встречается в том месте, в котором в ПСП стоит открывающаяся скобка (так как закрывающаяся скобка номер k всегда идет после открывающейся скобки номер k), а значит первое вхождение числа k будет позже, чем первые вхождений чисел $1 \dots (k-1)$. Результат — полупалиндром в силу симметричности ПСП: если на i-ом ($i \le n$) месте стоит k-ая открывающаяся скобочка, то на $n-i+1$ будет стоять закрывающаяся, и ее номер будет равен $n-k+1$, и аналогично с закрывающейся.

f — инъекция: для любых двух различных ПСП в месте их первого отличия в образе будут стоять разные числа.

Покажем, что f — сюръекция. Рассмотрим полупалиндром w на n буквах. Построим следующую скобочную последовательность s: если w_i встречается в w первый раз, $s_i =$ (, иначе $s_i =$). s_i — ПСП, так как баланс равен нулю и на любом префиксе баланс неотрицателен (так как каждое число мы встретим первый раз хотя бы столько же раз, сколько второй раз). s_i — симметричная ПСП, так как если в полупалиндроме w_i встречается первый раз, то w_{n-i+1} — второй, и наоборот. Так как $f(s) = w$ получаем, что f — сюръекция.

Так как f — инъекция и сюръекция, то f — биекция. □

Этой биекции можно придать больше смысла, введя для полупалиндрома аналогию открывающихся и закрывающихся скобок.

Определение 55. Пусть p — полупалиндром порядка n. Тогда s — префикс полупалиндрома p, если s является префиксом p как строки и длина s меньше n.

Предложение 56. *Пусть s — префикс полупалиндрома. Тогда s можно продолжить*

- *единственным способом, если каждая буква встречается в s дважды. Этот способ — новая буква (следующая по возрастанию после наибольшей среди встречающихся). Эта ситуация соответствует префиксу ПСП с нулевым балансом, единственный способ продолжить который — открывающая скобка.*

- *двумя способами, если существует буква, которая встречается в s один раз. Первый способ — новая буква (следующая по возрастанию после наибольшей среди встречающихся). Второй способ — наименьшая буква префикса, которая встречалась один раз. Эта ситуация соответствует префиксу ПСП с ненулевым балансом, первый способ продолжить который — открывающая скобка, второй — закрывающая.*

Пример 57. 112 можно продолжить или 2, или 3, а 1212 только 3. При этом, 123 нельзя продолжить ни 2, ни 3, а только 1 или 4.

Доказательство. Очевидно, мы не можем продолжить числом, большим чем минимальное, которое еще не встречалось — таким образом, мы нарушим возрастающий порядок. Покажем, что второе вхождение числа m, $m > k$ в полупалиндроме невозможно раньше, чем второе вхождение числа k. Пусть k — минимальное число, которое встречается один раз до того момента, как какое-то m встречается два раза. У нас может быть два случая: k второй раз встретился в первой половине полупалиндрома, или во второй. В первом случае у нас нарушится возрастающий порядок во второй половине полупалиндрома: в первой у нас стоит $k...m.m..k..$, значит во второй будет стоять $..(n-k+1)..(n-m+1).(n-m+1)...(n-k+1)$, но $n-k+1$ больше, чем $n-m+1$. Во втором случае k второй раз встречается во второй половине полупалиндрома, значит $n-k+1$ встречается в первой. Но это означает, что и все числа меньше $n-k+1$ тоже встречаются в первой, то есть там встречается $n-k+1$ различное число. При этом k — первое число, которое встречается один раз, значит все числа от 1 до $k-1$ встречаются два раза, а еще m встречается два раза. В n мест не помещается $n-k+1+k-1+1 = n+1$ число. Получается противоречие, доказывающее требуемое. □

Лемма 58. *Множество префиксов ПСП длины n биективно множеству симметричных ПСП длины $2n$.*

Доказательство. Пусть p — префикс ПСП. Если его отразить (развернуть и заменить открывающиеся скобки на закрывающиеся и наоборот) и присоединить к p, получится симметричная ПСП. □

Предложение 59. *Количество полупалиндромов на n буквах есть $SP_n = \binom{n}{\lfloor \frac{n}{2} \rfloor}$.*

Доказательство. Количество полупалиндромов на n буквах равно количеству префиксов ПСП длины n по лемме 54.

Пусть P_n^k — количество префиксов ПСП длины n, в которых меньше либо равно k закрывающихся скобок. Если $k > \lfloor \frac{n}{2} \rfloor$, то $P_n^k = P_n^{k-1}$. Действительно, ведь существует ноль префиксов ПСП, в которых больше половины закрывающихся скобок. Иначе, если $k \le \lfloor \frac{n}{2} \rfloor$, то $P_n^k = P_{n-1}^k + P_{n-1}^{k-1}$, так как последний символ может быть или открывающейся скобкой, или закрывающейся. Из-за рекуррентной формулы и начальных условий замечаем, что $P_n^{\lfloor \frac{n}{2} \rfloor} = \binom{n}{\lfloor \frac{n}{2} \rfloor}$, а $P_n^{\lfloor \frac{n}{2} \rfloor}$ — количество всех префиксов ПСП длины n.

Данная последовательность содержится в On-Line Encyclopedia of Integer Sequences[9] с индексом A001405. □

Предложение 60. *Количество неразложимых полупалиндромов на n буквах равно количеству сильно-неразложимых полупалиндромов на n буквах.*

Доказательство. В общем случае верно, что сильно-неразложимое 2-слово является неразложимым 2-словом. Надо показать, что неразложимый полупалиндром является сильно-неразложимым полупалиндромом.

Для того, чтобы это показать, покажем, что если у полупалиндрома есть 2-подслово, то у него есть и префикс, являющийся 2-словом. Обозначим данное 2-подслово как w. Если w префикс, то мы нашли искомый префикс. Если w находится целиком в первой половине, то рассмотрим u — префикс до начала w невключительно. Пакажем, что u — 2-слово. Пусть это не так, то есть какая-то буква a встречается в нем только один раз. Если a не встречается второй раз в u, значит второй раз она встречается после конца w (она не может встречаться в w, так как в нем каждая буква встречается два раза). Но $w_1 > a$, при этом встречается второй раз раньше, чем a встречается второй раз, что противоречит предложению 56, а значит u — 2-слово. Если w лежит целиком во второй половине, то симметрично ему есть 2-слово, которое целиком лежит в первой половине, из-за симметричности полупалиндрома. Остается случай, если w пересекает середину. Тут есть два случая. Первый случай, если середина w не является серединой полупалиндрома. Тогда симметрично w тоже лежит 2-слово v. Так как середина w не является серединой полупалиндрома, в w есть часть, которая не лежит в v. Эта часть является 2-словом, так как та часть, которая лежит и в w, и в v является 2-словом (потому что все буквы, которые в ней находятся, находятся оба раза и в w, и в v). Эта часть лежит целиком или в первой половине, или во второй, значит мы перешли к уже рассмотренному случаю. Последний случай — середина w совпадает с серединой полупалиндрома. w разделяется серединой на w_1 и w_2. Если в w_1 какая-то буква встречается два раза, значит префикс до w невключительно является 2-словом аналогично случаю, разобранному ранее. Если в w_1 каждая буква встречается по одному разу, значит w_1 длины m есть $k(k+1)\dots(k+m-1)$. Так как полупалиндром, w_2 тоже возрастает, и чтобы w было 2-словом, надо, чтобы $n-(k+m-1)+1=k$, то есть $n-m=2(k-1)$. До буквы k встречаются все $k-1$ меньшие буквы. Они стоят на $n-m$ местах (на оставшихся m местах первой половины стоит w_1). То есть $k-1$ букв стоят на $2(k-1)$ местах, значит каждая буква стоит два раза, значит это 2-слово. □

Предложение 61. *Количество сильно-неразложимых (просто неразложимых) полупалиндромов на n буквах есть $SPI_n = \binom{n-1}{\lfloor \frac{n-1}{2} \rfloor}$.*

Доказательство. Если есть префикс, являющийся 2-словом, длины больше n, то оставшийся суффикс тоже является 2-словом, и в силу симметричности полупалиндрома это значит, что есть и префикс, являющийся 2-словом, длины меньше либо равной n.

Рассмотрим биекцию между полупалиндромами и префиксами ПСП. Заметим, что если у полупалиндрома есть префикс длины k, являющийся 2-словом, то при биекции он перейдет в префикс ПСП, у которого баланс префикса длины k есть ноль. Получается неразложимые полупалиндромы переходят в префиксы ПСП, у которых никогда нет баланса ноль. То есть в скобочные последовательности, баланс на любом префиксе которых больше либо равен единице. Такие скобочные последовательности длины n — открывающаяся скобка конкатенированная с префиксом ПСП длины $n-1$, ведь префиксы ПСП — скобочные последовательности, баланс на любом префиксе которых больше либо равен нуля. Получается количество искомых объектов длины n равно количеству префиксов ПСП длины $n-1$, то есть $\binom{n-1}{\lfloor \frac{n-1}{2} \rfloor}$. □

Далее речь пойдет о гипотезе, что все полупалиндромы реализуемы.

Предложение 62. *Не у всех полупалиндромов сборное число равно единице.*

Доказательство. С помощью программы [10, `assembly_number_2.cpp`] был найден полупалиндром 112342534566 (рисунок 1), который имеет сборное число 2 и является самым коротким, наравне с 112345234566, полупалиндромом со сборным числом 2. □

Определение 63. Аналогично R_{min} (определение 32) определим R^P_{min} и R^{SP}_{min} — минимальное реализующее число для палиндромов и полупалиндромов.

Предложение 64. *В предположении, что $R_{min}(n) = 3n+2$, R^P_{min} совпадает с R_{min}.*

Доказательство. Серия примеров для R_{min} состоит из палиндромов, поэтому является серией примеров и для R^P_{min}. □

Рассмотрим серию 2-слов $\{K_i\}$. K_1 — 11, K_2 — 112342534566 (рисунок 1). K_n есть K_{n-1}, в конец которого добавлено (пусть $m = 5(n-2)$) $m+2, m+3, m+4, m+2, m+5, m+3, m+4, m+5, m+6, m+6$. На рисунке 2 — K_4

Предложение 65. *K_n — полупалиндром.*

Доказательство. Доказательство будем вести по индукции.

База: K_1 и K_2 являются полупалиндромами.

Переход: Пусть для всех $i < n$ K_i является полупалиндромом. Тогда K_n — тоже полупалиндром.

$n > 2$. В начале K_n записано 1123425345, в конце $m+2, m+3, m+4, m+2, m+5, m+3, m+4, m+5, m+6, m+6$. Заметим, что для $1 \le j \le 10$ $K_n[i] + K_n[2(5(n-1)+1) - i + 1] = m + 7 = 5(n-2)+7 = (5(n-1)+1)+1$, то есть $|K_n|+1$, что и нужно для полупалиндрома. Заметим, что то, что остается между, есть K_{n-2}, но начинающийся не с единицы, а с 6. Так как по предположению индукции K_{n-2} — полупалиндром, то и K_n — полупалиндром. □

Предложение 66. *K_n имеет сборное число, равное n.*

Доказательство. Доказательство будем вести по индукции.

База: $\mathrm{An}(K_1) = 1$ и $\mathrm{An}(K_2) = 2$.

Переход: Пусть для все $i < n$ $\mathrm{An}(K_i) = i$. Тогда $\mathrm{An}(K_n) = n$.

$\mathrm{An}(K_{n-1}) = n-1$.

Покажем, что $\mathrm{An}(K_n) \le n$. Действительно, гамильтоного множество полигональных путей, покрывающее весь граф, такого: возьмем минимальное гамильтоного множество полигональных путей для K_{n-1}, в нем $n-1$ элемент и добавим туда следующий путь $(m+6, m+5, m+2, m+3, m+4)$ (рисунок 3). Этот путь покрывает все вершины и является полигональным, а так же он покрывает только добавленные вершины, поэтому никак не пересекается с путями, которые уже были в множестве.

Покажем, что $\mathrm{An}(K_n) \geq n$. От противного, пусть это не так, то есть пусть $\mathrm{An}(K_n) \leq n-1$. 6 максимальных вершин образуют граф K_2, про который мы знаем, что $\mathrm{An}(K_2) = 2$. При этом существует максимум один путь, который покрывает и вершины из K_{n-1}, и добавленные, потому что единственное ребро, которое связывает эти два множества вершин есть ребро $(m+1, m+2)$. Значит если мы выкинем добавленные вершины, мы точно выкинем хотя бы один путь. Значит $\mathrm{An}(K_{n-1})$ будет меньше либо равен $n-2$, что противоречит предположению индукции и доказывает переход.

□

Предложение 67. *Для любого натурального n* $R^{SP}_{min}(n) \leq 5(n-1)+1$.

Доказательство. K_n содержит $5(n-1)+1$ вершин, является полупалиндромом (по предложению 65) и имеет сборное число, равное n (по предложению 66). □

Предложение 68. *С помощью программы[10] получена таблица 4: для каждого n найдено максимальное сборное число, достижимое на полупалиндромах на n буквах.*

Значения в таблице, вопреки ожиданиям, невозрастают: для $n = 6$ значение равно 2, а для $n = 7$ значение равно 1. Это место было дополнительно проверено вручную перебором 35 графов.

3.2 Практические

Была разработана библиотека[10] на языке C++[7] для работы с 2-словами и представлении их в виде сборных графов. Язык C++ был выбран из-за его быстродействия. Было решено не использовать объектно-ориентированное программирование, а избрать дизайн, ориентированный на данные[5]. Это позволяет библиотеке быть легко расширяемой и способствует производительности. Быстродействие так важно, потому что алгоритм поиска сборного числа [69] работает экспоненциально долго.

Библиотека реализует функции `is_double_occurrence_word`, `to_ascending_order`, `is_in_ascending_order`, `reverse`, `is_palindrome`, `equal_as_double_occur` `is_semi_palindrome`, `is_reducible`, `is_strongly_reducible`, `next_in_ascending_order`, `next_palindrome`, `next_semi_palindrome`, `assembly_number`, `minimal_realization_number_and_its_realization`, функциональность которых следует из названий.

Реализована функция `draw_as_graph`, использующая систему для визуализации графов Graphviz[8], которая изображает 2-слово в виде сборного графа (рисунок 1). На рисунке соблюден порядок ребер в каждой вершине, проходя каждый раз прямо, можно пройти по трансверсали. Синим обозначено минимальное гамильтоново множество полигональных путей.

Алгоритм 69. *Алгоритм для поиска сборного числа.*

Для поиска сборного числа необходимо перебрать все множества ребер, что все ребра в множестве попарно полигональны, проверить, что ребра не образуют циклов, вычислить количество путей, образованных ребрами, и вычислить минимум таких количеств по всем множествам. Пара ребер не полигональна тогда и только тогда, когда ребра этой пары идут в трансверсали не подряд. Таким образом, чтобы перебрать все множества ребер такие, что все ребра в множестве попарно полигональны, можно перебрать все множества ребер такие, что никакие два ребра не идут подряд. Опишем, как для фиксированного множества ребер E проверить, что ребра из E не образуют циклов и вычислить количество путей, образованных ребрами E. Структура “система непересекающихся множеств”[6] (СНМ) позволяет объединять два непересекающихся множества и проверять, лежат ли два элемента в одном множестве за $\mathcal{O}(\alpha(n))$, где $\alpha(n)$ — обратная функция Аккермана. Заведем такую структуру для одноэлементных множеств от $\{1\}$ до $\{n\}$, где множеству $\{i\}$ сопоставляется i-ая вершина. Будем перебирать ребра по порядку. Если очередное ребро соединяет вершины из разных множеств, объединим их. Иначе, оно соединяет вершины из одного множества, значит ребра E на вершинах этого множества образуют цикл, и множество E не подходит. По окончанию этого процесса каждое множество в СНМ будет связано ребрами E. Так как в множестве E нет пары неполигональных ребер, каждой вершине инцидентно или одно, или два ребра из E. Так как множество E не образует циклов, получается, что каждое множество в СНМ покрыто путем из E. Таким образом, чтобы узнать количество путей, образованных ребрами E, можно посчитать количество различных множеств в СНМ.

Пусть n — порядок Г. Интересующих нас множеств ребер $\mathcal{O}(2^{2n})$, для каждого множества необходимо выполнить $\mathcal{O}(n\alpha(n))$ операций. Таким образом, итоговая асимптотика алгоритма есть $\mathcal{O}(2^{2n}n\alpha(n))$.

Список литературы

[1] A. Angeleska, N. Jonoska, M. Saito DNA recombinations through assembly graphs // Discrete Applied Mathematics. - 2009. - №157. - С. 3020-3037.

[2] J. Burns, E. Dolzhenko, N. Jonoska, T. Muche, M. Saito Four-regular graphs with rigid vertices associated to DNA recombination // Discrete Applied Mathematics. - 2013. - №161. - С. 1378-1394.

[3] А. Э. Гутерман, Е. М. Крейнес, Н. В. Остроухова 2-слова: их графы и матрицы // Записки научных семинаров ПОМИ. - 2019. - №482. - С. 45-72.

[4] М.Вялый, В.Подольский, А.Рубцов, Д.Шварц, А.Шень Лекции по дискретной математике. - Изд. Дом ВШЭ, 2021. - 495 с.

[5] R. Fabian Data-oriented design. - 2018. - 307 с.

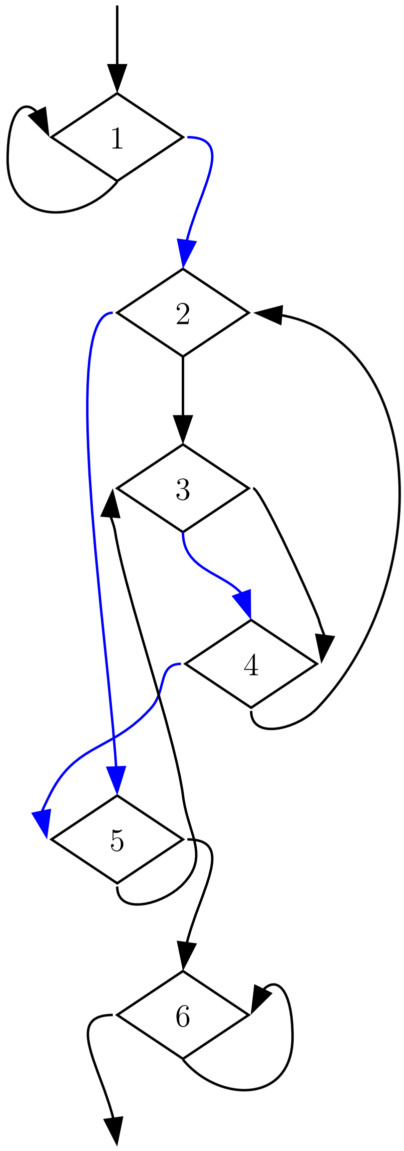
[6] Томас Кормен и др. Алгоритмы: построение и анализ. — 2-е изд. — М.: «Вильямс», 2006. — 1296 с.

[7] B. Stroustrup The C++ Programming Language. - 4-е изд. - Addison-Wesley Professional, 2013. - 1376 с.

[8] Emden R. Ganser, Stephen C. North An open graph visualization system and its applications to software engineering // Software: practice and experience. - 2000. - С. 1203-1233.

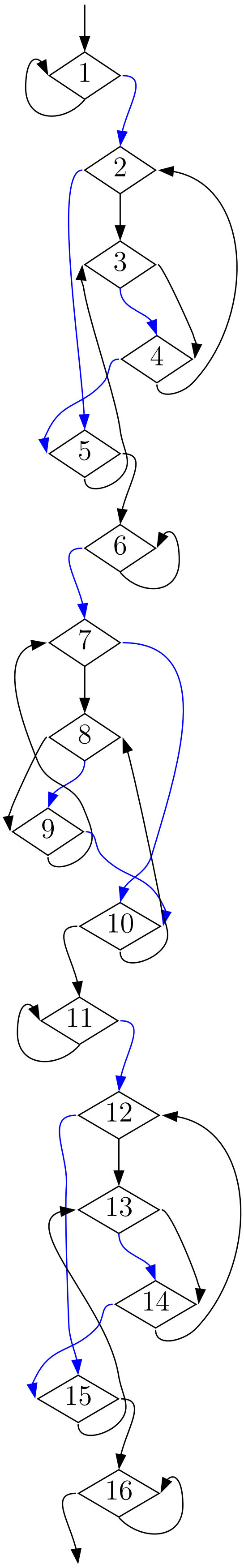
[9] The On-Line Encyclopedia of Integer Sequences `https://oeis.org`

[10] Репозиторий с исходным кодом `https://github.com/didedoshka/double_occurrence_words`



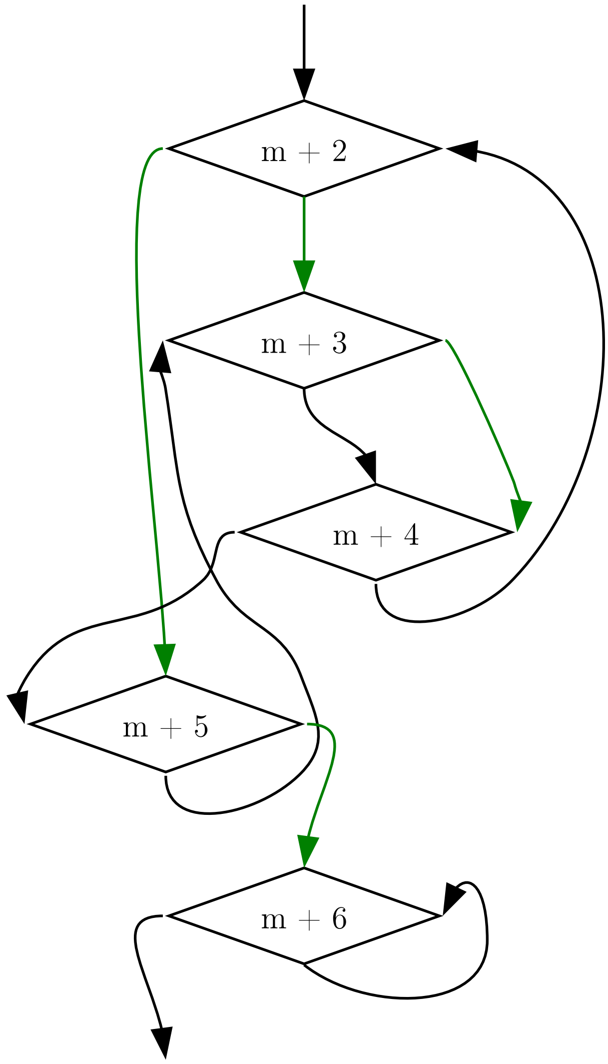
1 1 2 3 4 2 5 3 4 5 6 6
A double occurrence word
Already in ascending order
Palindrome: yes
Semi-palindrome: yes
Irreducible: no
Strongly-irreducible: no
Assembly number: 2
Unrealizable

Рис. 1: Результат работы программы.



1 1 2 3 4 2 5 3 4 5 6 6 7 8 9 7 10 8 9 10 11 11 12 13 14 12 15 13 14 15 16 16

Рис. 2: Пример полупалиндрома со сборным числом 4.



m + 2 m + 3 m + 4 m + 2 m + 5 m + 3 m + 4 m + 5 m + 6 m + 6

Рис. 3: Полигональный путь.

n	Все слова W_n	Палиндромы P_n	Полупалиндромы SP_n
1	1	1	1
2	3	3	2
3	15	7	3
4	105	25	6
5	945	81	10
6	10395	331	20
7	135135	1303	35
8	2027025	5937	70
9	34459425	26785	126
10	654729075	133651	252
11	13749310575	669351	462
12	316234143225	3609673	924
13	7905853580625	19674097	1716
14	213458046676875	113525595	3432
15	6190283353629375	664400311	6435
OEIS[9]	A001147	A047974	A00140

Таблица 1: Количество 2-слов, палиндромов и полупалиндромов.

n	Неразл. I_n	Неразл. палин. J_n	Неразл. полупалин. SPI_n
1	1	1	1
2	2	2	1
3	10	6	2
4	74	20	3
5	706	72	6
6	8162	290	10
7	110410	1198	20
8	1708394	5452	35
9	29752066	25176	70
10	576037442	125874	126
11	12277827850	637926	252
12	285764591114	3448708	462
13	7213364729026	18919048	924
14	196316804255522	109412210	1716
15	5731249477826890	642798510	3432
OEIS[9]	A000698	A195186	A001405

Таблица 2: Количество неразложимых 2-слов, неразложимых палиндромов и неразложимых полупалиндромов.

n	Сил.-нер. S_n	Сил.-нер. пал. T_n	Сил.-нер. полупал. SPI_n
1	1	1	1
2	1	1	1
3	4	2	2
4	27	7	3
5	248	22	6
6	2830	96	10
7	38232	380	20
8	593859	1853	35
9	10401712	8510	70
10	202601898	44940	126
11	4342263000	229836	252
12	101551822350	1296410	462
13	2573779506192	7211116	924
14	70282204726396	43096912	1716
15	2057490936366320	256874200	3432
OEIS[9]	A000699	A004300	A001405

Таблица 3: Количество сильно-неразложимых 2-слов, сильно-неразложимых палиндромов и сильно-неразложимых полупалиндромов.

n	Макс. An, достиж. на полупал. на n буквах
1	1
2	1
3	1
4	1
5	1
6	2
7	1
8	2
9	2
10	2
11	3
12	3
13	3
14	3
15	3

Таблица 4: Максимальное сборное число, достижимое на полупалиндромах на n буквах.