ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Образовательная программа «Прикладная математика и информатика»

Отчет об исследовательском проекте

на тему Изучение свойств сборных графов – палиндромов и полупалиндромов

(промежуточный, этап 1)

Выполнен студентом:

Группы #БПМИ231 Горохов Дмитрий Александрович

ФИО студента

Проверен руководителем проекта:

Максаев Артем Максимович, к. ф.-м. н.

ФИО, научная степень (если есть)

Доцент

Должность

Факультет компьютерных наук /
Департамент больших данных и
информационного поиска

Место работы (организация или департамент НИУ ВШЭ)

Москва 2025

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

[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.

[2] же 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.

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

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. 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. 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. 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. 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. *В простом сборном графе существует единственный класс эквивалентности трансверсалей.*

Лемма 16. *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. Обратное слово к слову $w = a_{i_1} \dots a_{i_k}$ (обозначение w^R) — $a_{i_k} \dots a_{i_1}$

Определение 19. Два 2-слова эквивалентны, если после переименования некоторых букв они или совпадают, либо являются обратными друг для друга.

Лемма 20. *Классы эквивалентности 2-слов находятся в биективном соответствии с классами изоморфизма простых сборных графов.*

Определение 21. 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 22. Композиция простых сборных графов — простой сборный граф.

Определение 23. 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.

Определение 24. Простой путь — путь, не содержащий какую-либо вершину дважды.

Определение 25. Множество попарно непересекающихся простых путей $\{\gamma_1, \dots, \gamma_k\}$ — гамильтоново, если их объединение содержит все вершины степени 4 графа Γ.

Определение 26. Полигональный путь — путь $(v_0, e_1, v_1, e_2, \dots, e_n, v_n)$, что e_i e_{i+1} — соседи для v_i для $i \in \{1, \dots, n-1\}$

Определение 27. Сборное число простого сборного графа Γ (обозначение $\mathrm{An}(\Gamma)$), определяется как $\mathrm{An}(\Gamma) = \min\{k|$ существует гамильтоново множество полигональных путей $\{\gamma_1, \dots, \gamma_k\}$ в $\Gamma\}$

Определение 28. Реализумый простой сборный граф — простой сборный граф, со сборным числом 1. Иначе — нереализуемый.

Лемма 29. *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 30. $\forall n \in \mathbb{N}$

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

Определение 31. 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 32. *Следующие свойства выполняются для* $R_{\min}$

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

Предложение 33. *Существует константа* N *такая, что для любого реализуемого 2-слова* w*, существует нереализуемое неразложимое* v*, что* $w \subset v$ *и* $|v| - |w| < N$

Предложение 34. *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 35. 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.

Лемма 36. *Мощность множества 2-слов на n буквах есть*

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

Определение 37. Палиндром — такое 2-слово, что его обратное (записанное в возрастающем порядке) равно ему.

Лемма 38. *Количество палиндромов на n буквах есть*

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

Лемма 39. *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 40. *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 41. Сильно-неразложимое 2-слово — такое 2-слово, что оно не содержит никакого собственного 2-подслова.

Лемма 42. *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 43. *Количество сильно-неразложимых палиндромов на n буквах есть*

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

3 Полученные результаты

3.1 Теоретические

Определение 44. 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 45. 1122 и 1212 — полупалиндромы, а 1221 — нет.

Предложение 46. *Количество полупалиндромов на n буквах есть $\binom{n}{\lfloor \frac{n}{2} \rfloor}$*

Предложение 47. *Количество неразложимых полупалиндромов на n буквах равно количеству сильно-неразложимых полупалиндромов на n буквах*

Предложение 48. *Не у всех полупалиндромов сборное число равно единице, в частности 112345234566 имеет сборное число 2 и является самым коротким, наравне с 112342534566, полупалиндромом со сборным числом 2.*

3.2 Практические

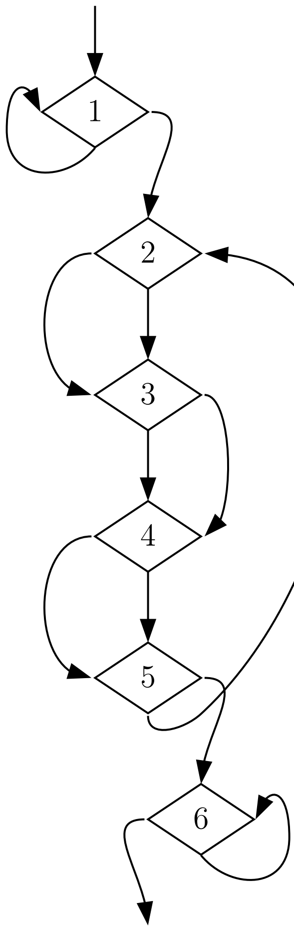
https://github.com/didedoshka/double_occurrence_words

Была разработана библиотека на языке C++ для работы с 2-словами и представлении их в виде сборных графов. Язык C++ был выбран из-за его быстродействия, так как алгоритм поиска сборного числа работает экспоненциально долго.

Библиотека реализует функции `is_double_occurrence_word`, `to_ascending_order`, `is_in_ascending_order`, `reverse`, `is_palindrome`, `equal_as_double_occurrence_words`, `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 названия.

Реализована функция `draw_as_graph`, которая изображает 2-слово в виде сборного графа (смотри 1). На рисунке соблюден порядок ребер в каждой вершине, идя каждый раз прямо, можно пройти по трансверсали.

Алгоритм для поиска сборного числа использует структуру данных "система непересекающихся множеств", что позволяет сократить ассимптотику до $\mathcal{O}(\binom{2n}{n} n\alpha(n))$, где $\alpha(n)$ — обратная функция Аккермана.



1 1 2 3 4 5 2 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

Список литературы

[1] M. S. Angela Angeleska, Nataša Jonoska. Dna recombination through assembly graphs. *Discrete Applied Mathematics*, 2009.

[2] N. J. T. M. M. S. Jonathan Burns, Egor Dolzhenko. Four-regular graphs with rigid vertices associated to dna recombination. *Discrete Applied Mathematics*, 2013.