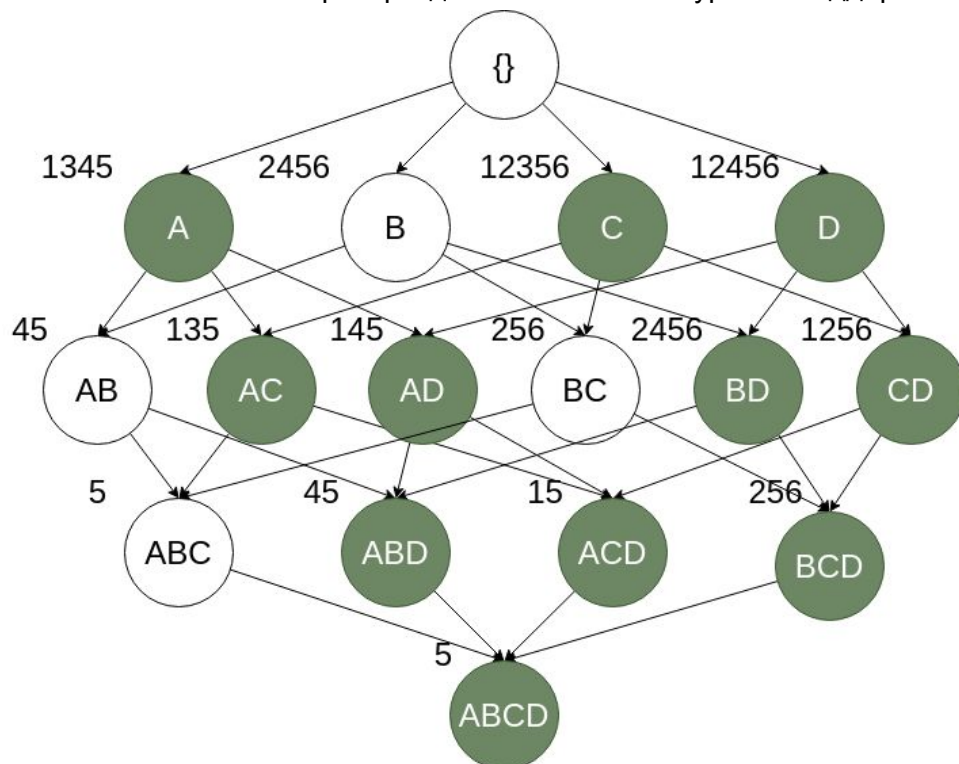


Задание №1

Дан набор данных:

Tid	Itemset
1	<i>ACD</i>
2	<i>BCD</i>
3	<i>AC</i>
4	<i>ABD</i>
5	<i>ABCD</i>
6	<i>BCD</i>

Найдите все минимальные генераторы для минимального уровня поддержки = 1.

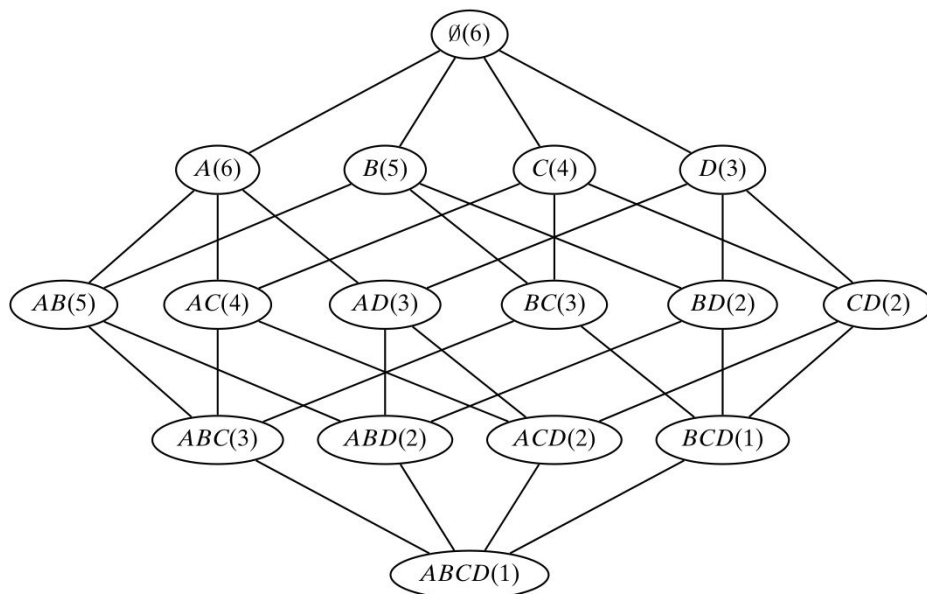


В таблице C - закрытые последовательности, G - минимальные генераторы.

tidset	C	G
1345	A	A
12356	C	C
12456	D	D
135	AC	AC
145	AD	AD
2456	BD	B
1256	CD	CD
45	ABD	AB
15	ACD	ACD
256	BCD	BC
5	ABCD	ABC

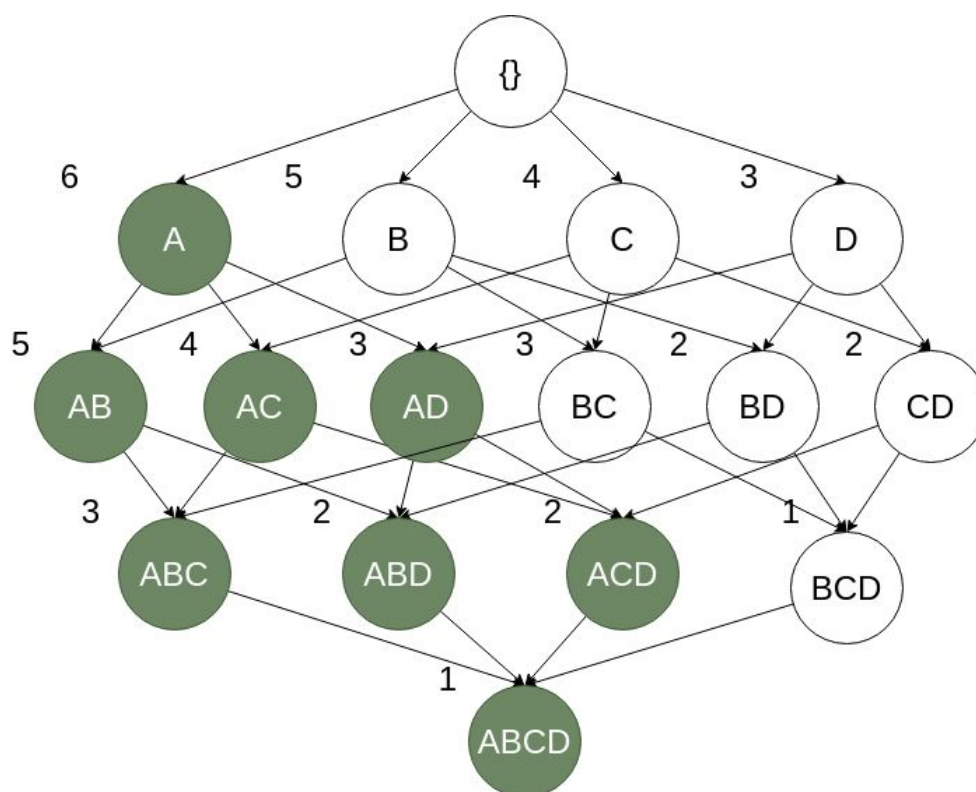
Задание №2

Дана решетка наборов и их частоты.



Выполните следующие задание:

- A. Выпишите список всех закрытых наборов (closed itemsets)



C
A
AB
AC
AD
ABC
ABD
ACD
ABCD

В. Является ли набор BCD выводимым? Является ли набор ABCD выводимым?
Какие границы их поддержки?.

а. BCD:

Нижняя граница уровня поддержки:

$$\text{sup}(\text{BCD}) \geq 0$$

$$\geq \text{sup}(\text{BC}) + \text{sup}(\text{BD}) - \text{sup}(\text{B}) = 3 + 2 - 5 = 0$$

$$\geq \text{sup}(\text{BC}) + \text{sup}(\text{CD}) - \text{sup}(\text{C}) = 3 + 2 - 4 = 1$$

$$\geq \text{sup}(\text{BD}) + \text{sup}(\text{CD}) - \text{sup}(\text{D}) = 2 + 2 - 3 = 1$$

$$LB(BCD) = \{0; 1\}$$

$$\max\{LB(ACD)\} = 1$$

Верхняя граница уровня поддержки:

$$\begin{aligned} \sup(BCD) &\leq \sup(BC) = 3 \\ &\leq \sup(BD) = 2 \\ &\leq \sup(CD) = 2 \\ &\leq \sup(BC) + \sup(BD) + \sup(CD) - \sup(B) - \sup(C) - \sup(D) + \sup(\{\}) = \\ &= 3 + 2 + 2 - 5 - 4 - 3 + 6 = 1 \end{aligned}$$

$$UB(BCD) = \{1; 2; 3\}$$

$$\min\{UB(ACD)\} = 1$$

$$\max\{LB(ACD)\} = \min\{UB(ACD)\} \Rightarrow \mathbf{BCD \text{ выводим}}$$

b. ABCD

Нижняя граница уровня поддержки:

$$\begin{aligned} \sup(ABCD) &\geq 0 \\ &\geq \sup(ABC) + \sup(ABD) - \sup(AB) = 0 \\ &\geq \sup(ABC) + \sup(ACD) - \sup(AC) = 1 \\ &\geq \sup(ABD) + \sup(ACD) - \sup(AD) = 1 \\ &\geq \sup(ABC) + \sup(BCD) - \sup(BC) = 1 \\ &\dots \\ &\geq \sup(ABC) + \sup(ABD) + \sup(ACD) + \sup(BCD) - \\ &\quad - \sup(AB) - \sup(AC) - \sup(AD) - \sup(BC) - \sup(BD) - \sup(CD) + \\ &\quad + \sup(A) + \sup(B) + \sup(C) + \sup(D) - \sup(\{\}) = 1 \end{aligned}$$

$$LB(ABCD) = \{0; 1\}$$

$$\max\{LB(ABCD)\} = 1$$

Верхняя граница уровня поддержки:

$$\begin{aligned} \sup(ABCD) &\leq \sup(ABC) = 3 \\ &\leq \sup(ACD) = 2 \\ &\leq \sup(ABD) = 2 \\ &\leq \sup(BCD) = 1 \\ &\leq \sup(ABC) + \sup(ABD) + \sup(ACD) - \sup(AB) - \sup(AC) - \sup(AD) + \\ &\quad + \sup(A) = 1 \\ &\dots \end{aligned}$$

$$UB(ABCD) = \{1; 2; 3\}$$

$$\min\{UB(ABCD)\} = 1$$

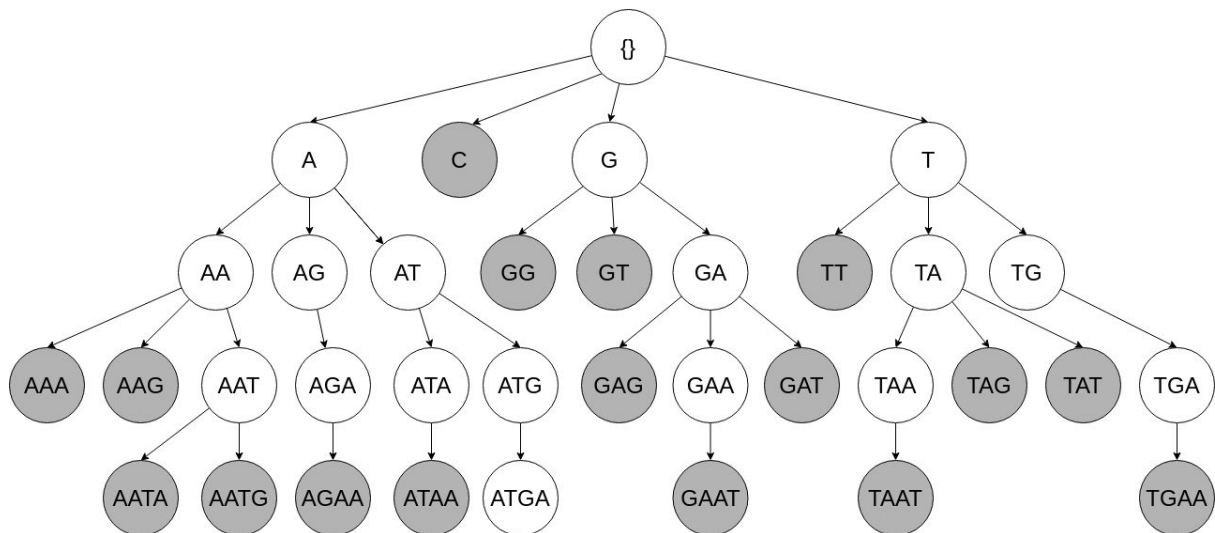
$$\max\{LB(ABCD)\} = \min\{UB(ABCD)\} \Rightarrow \mathbf{ABCD \text{ выводим}}$$

Задание №3

Даны последовательности

Id	Sequence
s_1	AATACAAGAAC
s_2	GTATGGTGAT
s_3	AACATGGCCAA
s_4	AAGCGTGGTCAA

А. Найдите все подпоследовательности в минимальном уровне поддержки = 4



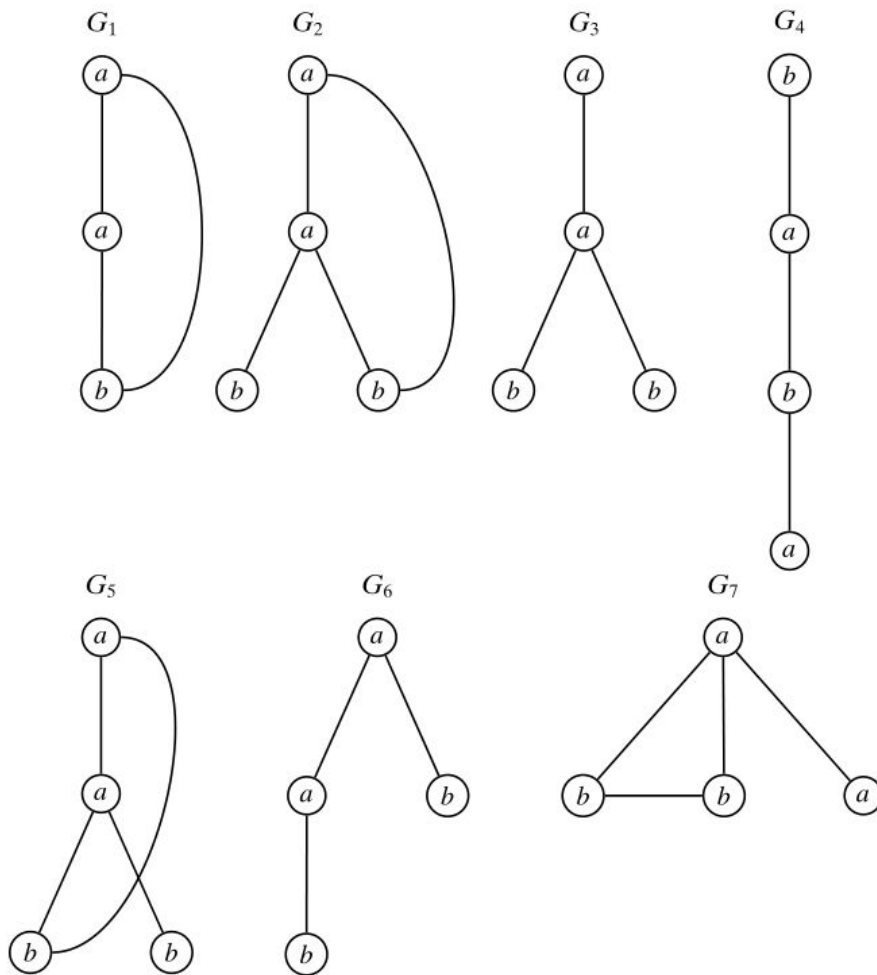
Белые ячейки - последовательности с уровнем поддержки 4, серые с уровнем поддержки меньше 4.

В. Для алфавита {A,C,G,T} посчитайте, сколько всего может быть разных последовательностей длины k

$$x = 4^k$$

Задание №4

Дан набор графов. Разделите их на изоморфные группы



- 1) G_2, G_5 изоморфны: G_1, G_3, G_4, G_6 изоморфные подграфы
- 2) G_7 : G_3 изоморфный подграф