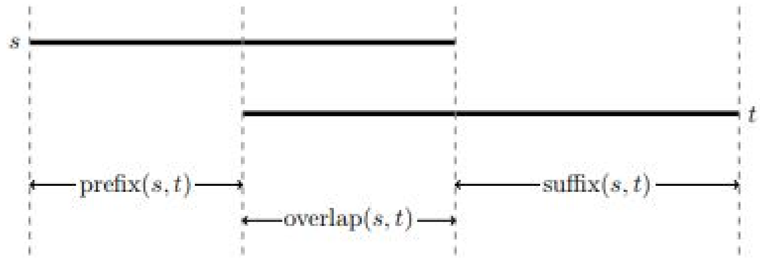
Приближенное построение наименьшей общей надстроки с помощью графа де Брёйля

Общая посlановка задачи SCS: дан набор строк $\mathcal{S}$ и требуется найти самую короткую строку, которая содержит каждую строку из $\mathcal{S}$. Сама задача NP-полна, но известны прибиженные алгоритмы, такие как 2, $\frac{2}{3}$ приближение, которое строится следующим образом: Строим префикс-граф (вершины - $\mathcal{S}$, ребро из s_1 в s_2 имеет вес $prefix(s_1, s_2)$), на нем оптимальное покрытие циклами. Затем в оверлэп-графе (вершины - эти циклы, ребра весом в $overlap(C_1, C_2)$) находим $\frac{2}{3}$ приближение задачи о максимальном коммивояжере.

$$\mathrm{prefix}(s,t) \circ \mathrm{overlap}(s,t) = s, \mathrm{overlap}(s,t) \circ \mathrm{suffix}(s,t) = t\,.$$



$$\mathrm{overlap}(\mathtt{ABACBA}, \mathtt{BABCA}) = \mathtt{BA},\ \mathrm{prefix}(\mathtt{ABACBA}, \mathtt{BABCA}) = \mathtt{ABAC}\,.$$

Задача r-SCS — любая строка из $\mathcal{S}$ имеет длину r.

Граф де Брёйля для $\mathcal{S}$ определяется следующим образом: каждая строка s из $\mathcal{S}$ представляется как ребро $prefix(s) \to suffix(s)$, где $prefix(s)$— s без последней буквы, а суффикс — без первой.

2-SCS решается в графе де Брёйля за полиномиальное время.

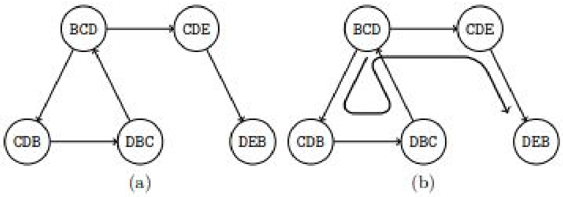


Рис. 1: (a)—Граф де Брёйля для $\mathcal{S} = \{BCDB, CDBC, DBCD, BCDE, CDEB\}$. (b) — эйлеров путь в сильносвязном графе де Брёйля дает наименьшую надстроку для $\mathcal{S}$ — $BCDBCDEB$

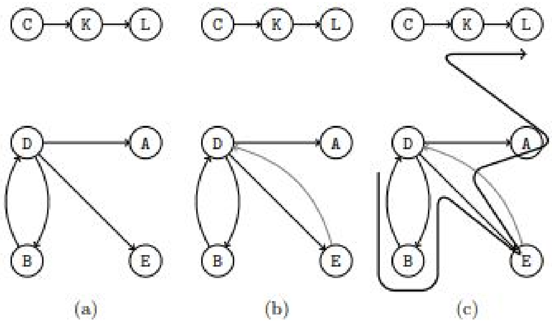


Рис. 2: (a) — граф де Брёйля для $\mathcal{S} = \{CK, KL, DA, DB, DE, BD\}$. (b) — добавление ребер, чтобы каждая компонента имела эйлеров путь. (c) — путь по всем компонентам в любом порядке дает наименьшую надстроку для $\mathcal{S}$ — $DBDEDACKL$

Algorithm 1 $(r^2 + r - 4)/(4r - 6)$-приближенный алгоритм для r-SCS

Вход: $\mathcal{S} = \{s_1, \ldots, s_n\} \subseteq \Sigma^r$

1: π — $\frac{2}{3}$-приближение максимального коммивояжера в overlap-графе от $\mathcal{S}$.

2: $\mathcal{S}' = \{s'_1, \ldots, s'_n\} \subseteq \Sigma_1^r$ — множество 2-строк над алфавитом $\Sigma_1 = \Sigma^{r-1}$. Где $s'_i = prefix(s_i)suffix(s_i)$

3: π_1 — 2-SCS для $\mathcal{S}'$

return $min(\pi, \pi_1)$