Autumn Festa 2026

Shortest Common Superstringの近似アルゴリズムについて

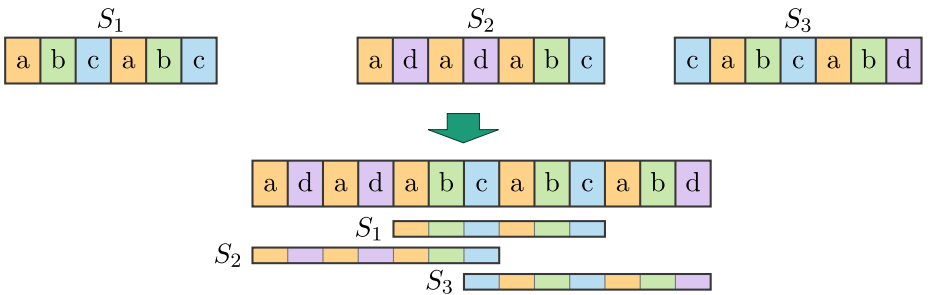
柴田 紘希（九州大学）

発表スライド

今回扱う問題

最短超文字列問題 (Shortest Common Superstring, SCS)

- **入力:** 文字列集合 $\mathcal{S} = \{S_1, \ldots, S_m\}$
- **出力:** $\mathcal{S}$ の**各文字列を部分文字列として含む**最短の文字列

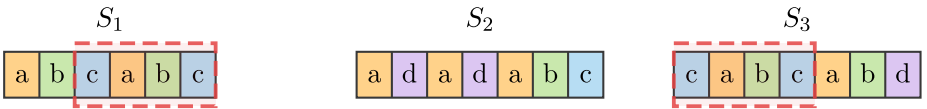


この問題は NP 困難で、（P $\neq$ NP なら）PTAS も存在しない

[Gallant et al. 1980, Blum et al. 1991]

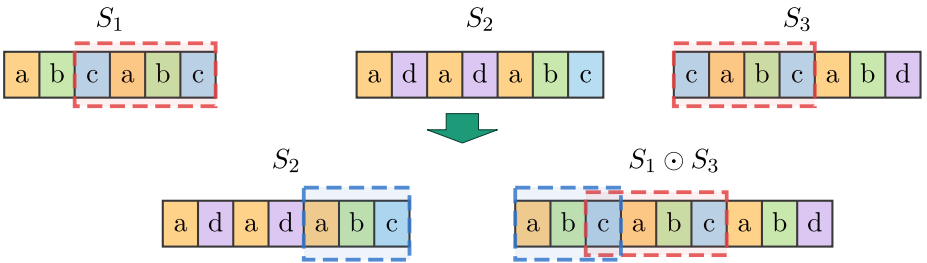
自然な貪欲アルゴリズム

GREEDY: **Overlap 部分の長さ**が最大な文字列同士を貪欲に結合するアルゴリズム



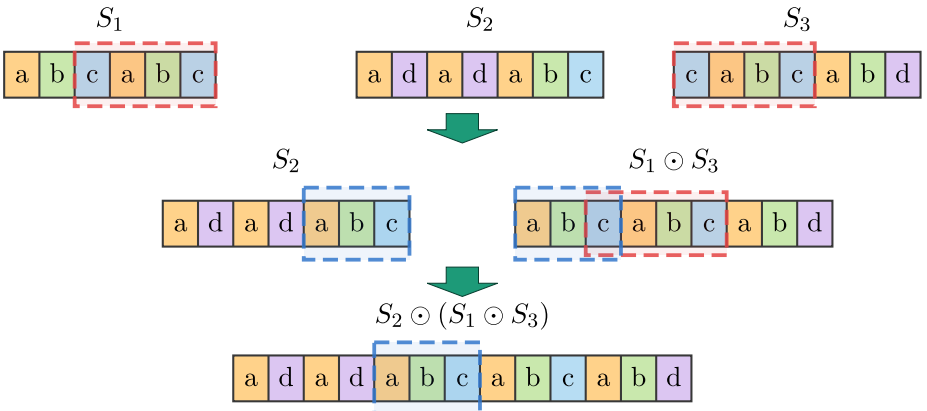
自然な貪欲アルゴリズム

GREEDY: **Overlap 部分の長さ**が最大な文字列同士を貪欲に結合するアルゴリズム



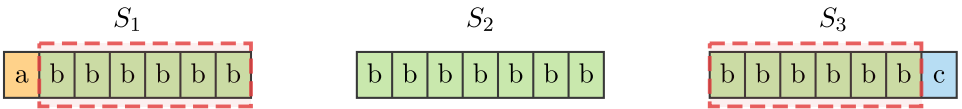
自然な貪欲アルゴリズム

GREEDY: **Overlap 部分の長さ**が最大な文字列同士を貪欲に結合するアルゴリズム



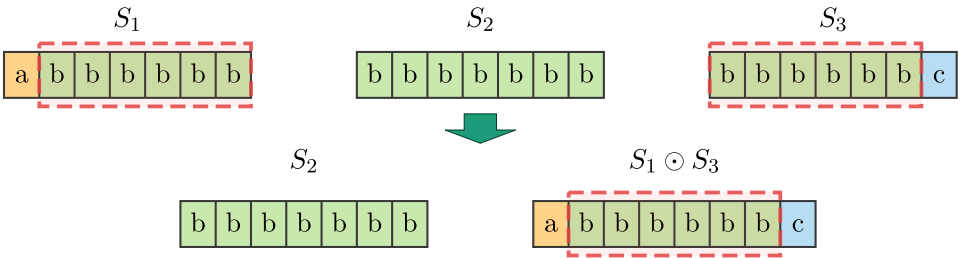
GREEDYの自然な近似比下界

以下のような入力では、GREEDY における**ある悪い実行の**近似比が 2 に漸近する



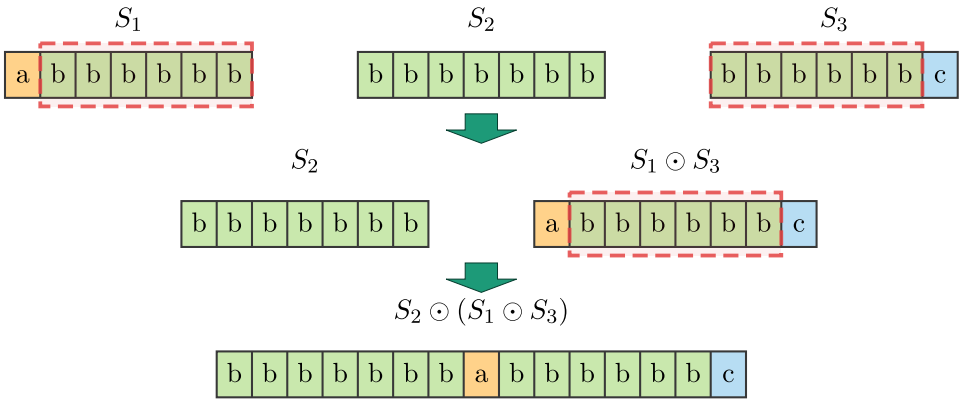
GREEDYの自然な近似比下界

以下のような入力では、GREEDY における**ある悪い実行の**近似比が 2 に漸近する



GREEDYの自然な近似比下界

以下のような入力では、GREEDY における**ある悪い実行の**近似比が 2 に漸近する



主結果

予想: Greedy Superstring Conjecture [Tarhio and Ukkonen, 1988]

GREEDY の近似比は 2 である。

Vazirani の “Approximation Algorithms” にも載っている、
約 40 年来の未解決予想

主結果

予想: Greedy Superstring Conjecture [Tarhio and Ukkonen, 1988]

GREEDY の近似比は 2 である。

Vazirani の “Approximation Algorithms” にも載っている、
約 40 年来の未解決予想

主結果: GREEDY の近似比下界

GREEDY の近似比は $12/5 = 2.4$ 以上である。

長年の未解決予想であった **Greedy Conjecture を否定的に解決**！

今回の発表内容

3 ステップ構成で紹介:

1. **近似比下界 $13/6 \geq 2.16$: 論文の結果を Simplify したもの**
2. **近似比下界 $9/4 = 2.25$: 初期論文の結果**
   - 主な改善点: cycle を 2 つ→ 3 つに更新
3. **近似比下界 $12/5 = 2.4$: 最新論文の結果（概略のみ）**
   - 主な改善点: 巡回文字列構成からの脱却

目次

1. **近似比下界** $13/6 \geq 2.16$

2. **近似比下界** $9/4 = 2.25$

3. **近似比下界** $12/5 = 2.4$ **（概略）**

今回の発表内容

3 ステップ構成で紹介:

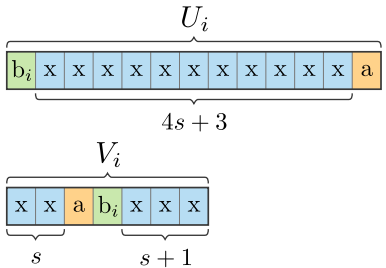
1. **近似比下界 $13/6 \geq 2.16$: 論文の結果を Simplify したもの**
2. 近似比下界 $9/4 = 2.25$: 初期論文の結果
   - 主な改善点: cycle を 2 つ→ 3 つに更新
3. 近似比下界 $12/5 = 2.4$: 最新論文の結果（概略のみ）
   - 主な改善点: 巡回文字列構成からの脱却

文字列集合 $\mathcal{S}$ の構成 (1/2)

2 つのパラメータから文字列集合 $\mathcal{S}$ を構成する:

- $k = 4s + 5 \geq 9$ (文字列長を表す)
- $t \geq 1$ (文字列の個数を表す)

アルファベット $\Sigma = \{\mathrm{a}, \mathrm{x}, \mathrm{b}_1, \ldots, \mathrm{b}_t\}$ 上の $2t$ 個の文字列 U_i, V_i $(1 \leq i \leq t)$ を以下のように定義し、そこから巡回文字列 U_i^∞, V_i^∞ を得る

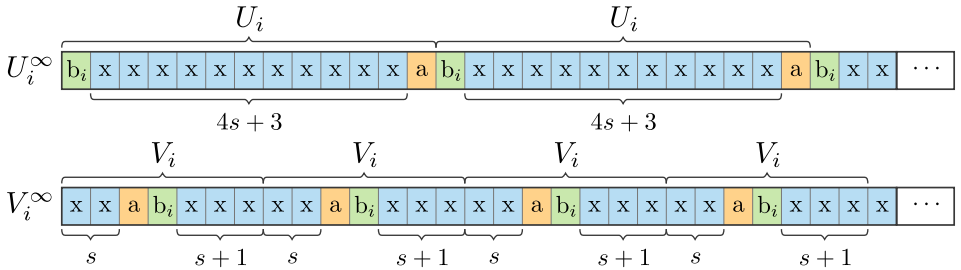


文字列集合 $\mathcal{S}$ の構成 (1/2)

2 つのパラメータから文字列集合 $\mathcal{S}$ を構成する:

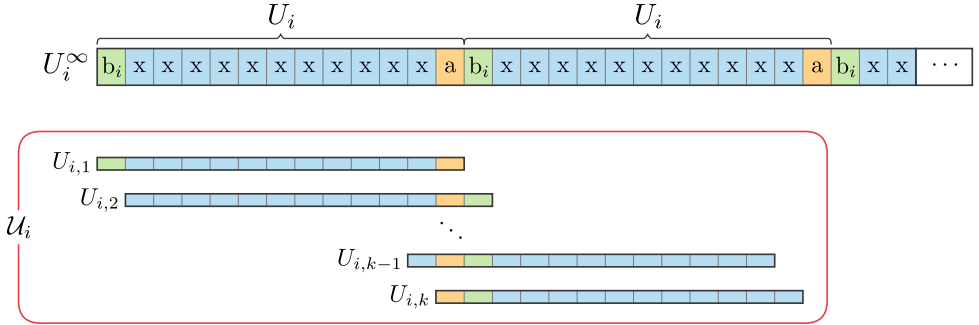
- $k = 4s + 5 \geq 9$ (文字列長を表す)
- $t \geq 1$ (文字列の個数を表す)

アルファベット $\Sigma = \{\mathrm{a}, \mathrm{x}, \mathrm{b}_1, \ldots, \mathrm{b}_t\}$ 上の $2t$ 個の文字列 U_i, V_i $(1 \leq i \leq t)$ を以下のように定義し、そこから巡回文字列 U_i^∞, V_i^∞ を得る



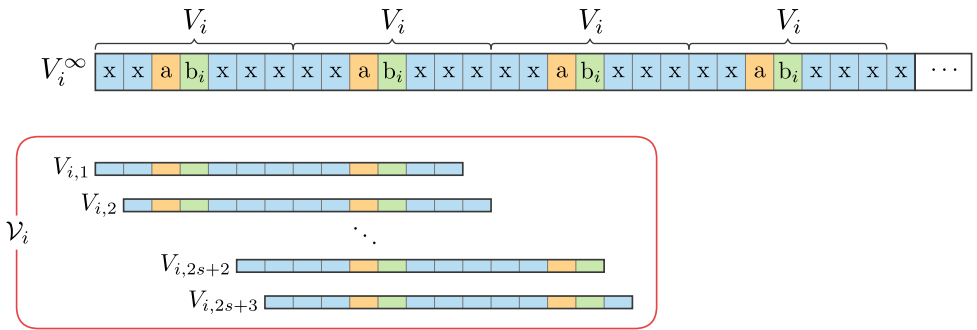
文字列集合 $\mathcal{S}$ の構成 (2/2)

各巡回文字列から長さ $k = 4s + 5 = |U_i|$ の部分文字列を切り出し、これらを入力文字列集合 $\mathcal{S} = (\mathcal{U}_1 \sqcup \cdots \sqcup \mathcal{U}_t) \sqcup (\mathcal{V}_1 \sqcup \cdots \sqcup \mathcal{V}_t)$ とする



文字列集合 $\mathcal{S}$ の構成 (2/2)

各巡回文字列から長さ $k = 4s + 5 = |U_i|$ の部分文字列を切り出し、これらを入力文字列集合 $\mathcal{S} = (\mathcal{U}_1 \sqcup \cdots \sqcup \mathcal{U}_t) \sqcup (\mathcal{V}_1 \sqcup \cdots \sqcup \mathcal{V}_t)$ とする



近似比下界の証明戦略

証明の戦略

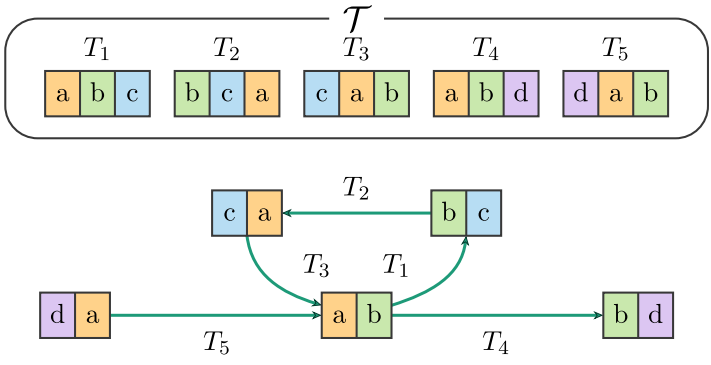
1. **上界:** 全ての文字列を長さ $k-1$ の overlap で結合できるような最適解が存在
   - **Rauzy Graph** を用いて構成

2. **下界:** 結合順序を恣意的に選び、出力サイズが大きくなるような GREEDY の解を構成

文字列集合の Rauzy Graph 表現

Rauzy Graph: 各文字列の長さが k の文字列集合 $\mathcal{T}$ から以下のように定義されるグラフ

- **頂点:** $\mathcal{T}$ に現れる長さ $k-1$ の部分文字列
- **辺:** $\mathcal{T}$ 中の各文字列
  - 長さ $k-1$ の接頭辞から接尾辞への辺になっている

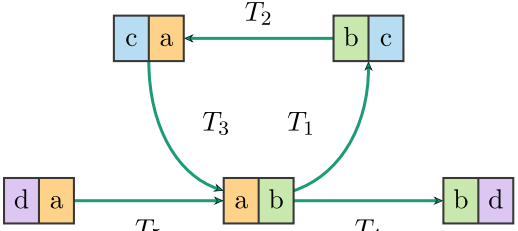


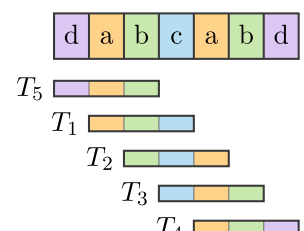
Rauzy Graph のオイラー路と SCS

補題: Rauzy Graph のオイラー路と SCS

$\mathcal{T}$ の Rauzy Graph がオイラー路を持つならば、$\mathcal{T}$ に対する長さ $|\mathcal{T}| + k - 1$ の Common Superstring が存在する。

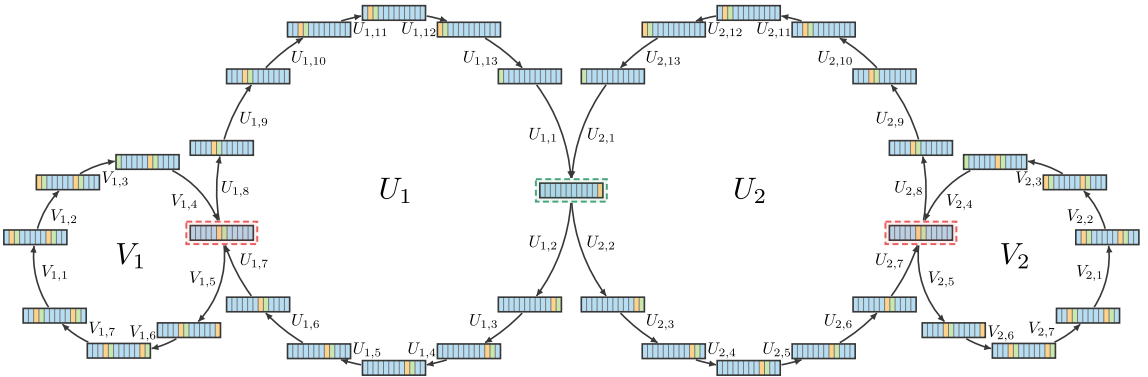
$\mathcal{T}$ の Common Superstring は長さ k の文字列 $|\mathcal{T}|$ 種類を含む必要があるため、長さ $|\mathcal{T}| + k - 1$ は最適で、これより短くなることはない





文字列集合 $\mathcal{S}$ の Rauzy Graph

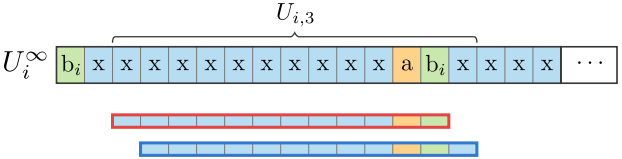
事実: $\mathcal{S}$ の Rauzy Graph は以下のようになり、オイラー路を持つ

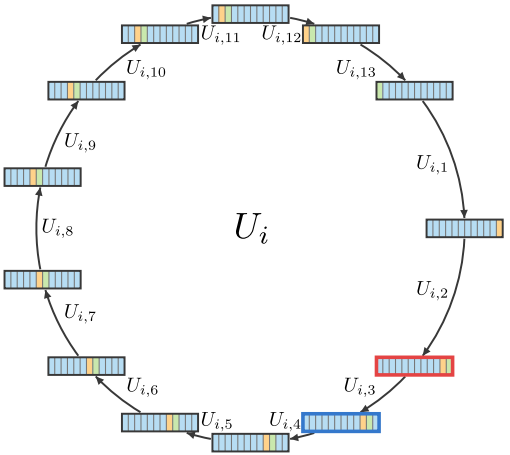


Rauzy Graph におけるサイクル

各文字列 $U_{i,j}$ は U_i^∞ の長さ k の部分文字列なため、辺集合 $\mathcal{U}_i$ はサイクルを成す

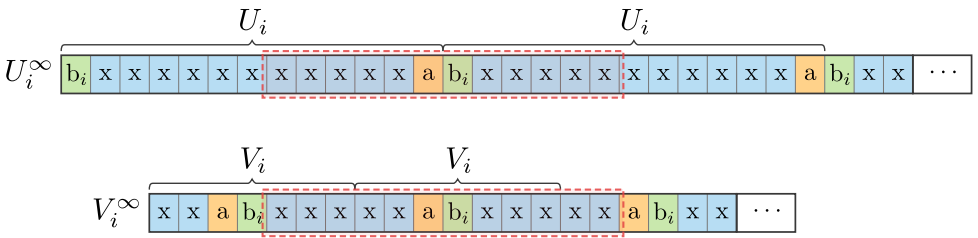
- V_i に関する部分も同様





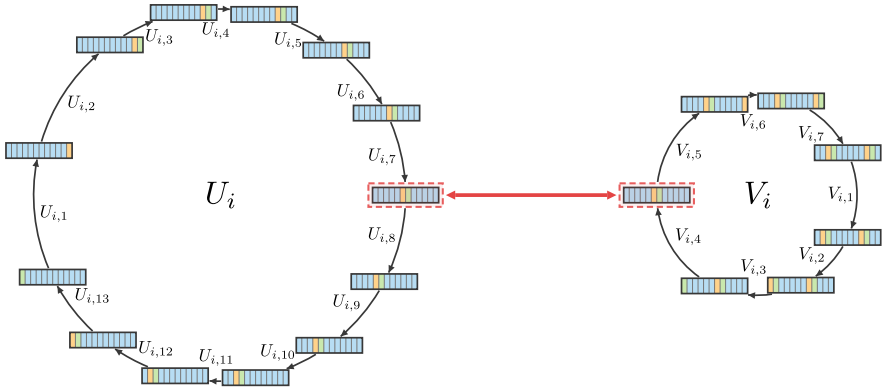
共通頂点によるサイクルの結合 (1/2)

(U_i^∞, V_i^∞) は長さ $k-1$ の共通部分文字列を持つ
→ U_i と V_i のサイクルが結合できる



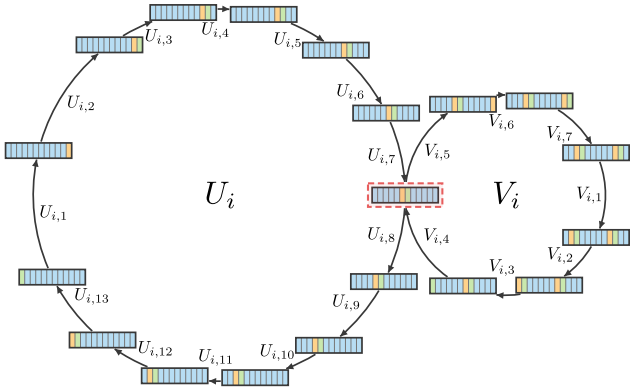
共通頂点によるサイクルの結合 (1/2)

$(U_i^{\infty}, V_i^{\infty})$ は長さ $k-1$ の共通部分文字列を持つ
→ U_i と V_i のサイクルが結合できる



共通頂点によるサイクルの結合 (1/2)

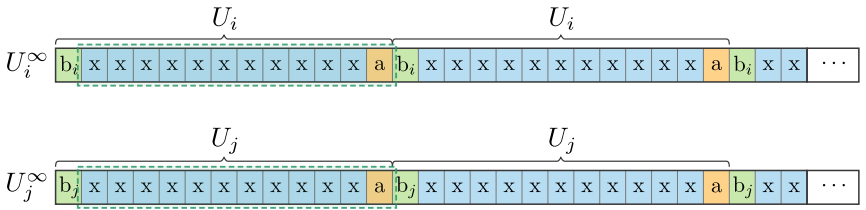
$(U_i^{\infty}, V_i^{\infty})$ は長さ $k-1$ の共通部分文字列を持つ
→ U_i と V_i のサイクルが結合できる



共通頂点によるサイクルの結合 (2/2)

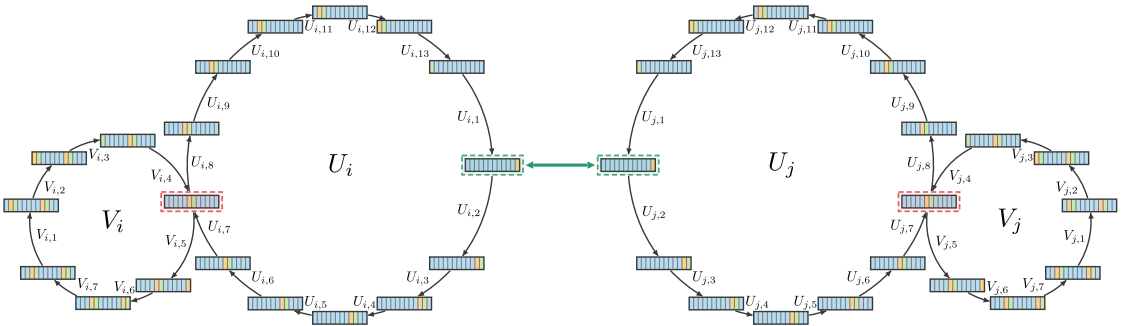
(U_i^∞, U_j^∞) は長さ $k-1$ の共通部分文字列を持つ

→ i, j それぞれのグラフが結合できる



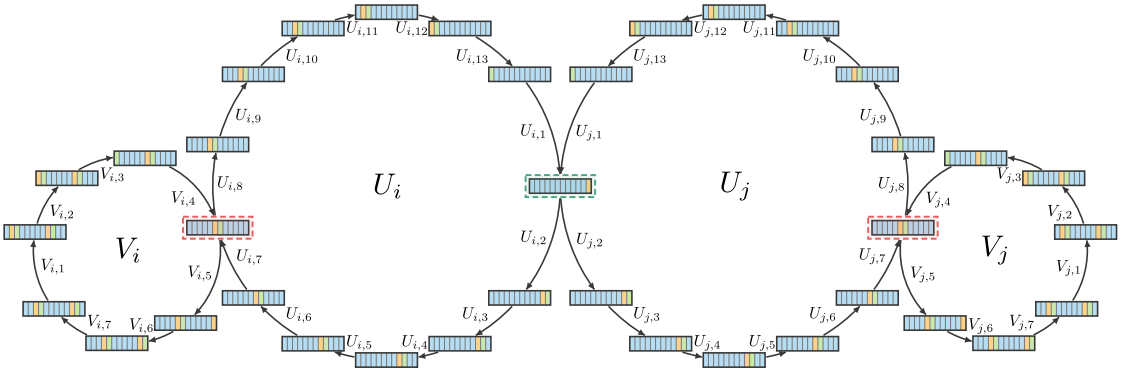
共通頂点によるサイクルの結合 (2/2)

$(U_i^{\infty}, U_j^{\infty})$ は長さ $k-1$ の共通部分文字列を持つ
→ i, j それぞれのグラフが結合できる



共通頂点によるサイクルの結合 (2/2)

$(U_i^{\infty}, U_j^{\infty})$ は長さ $k-1$ の共通部分文字列を持つ
→ i, j それぞれのグラフが結合できる

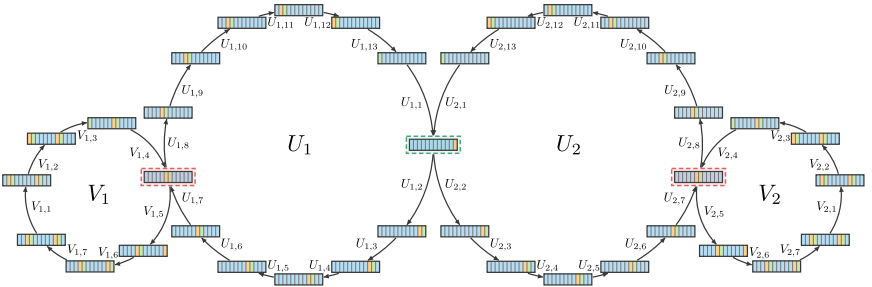


SCS 長の上界

$\mathcal{S}$ の Rauzy Graph がオイラー路を持つため、以下が成り立つ

定理 1: Shortest Common Superstring 長の上界

$\mathcal{S}$ は長さ $|\mathcal{S}| + k - 1$ の Common Superstring を持つ。



方針の概略

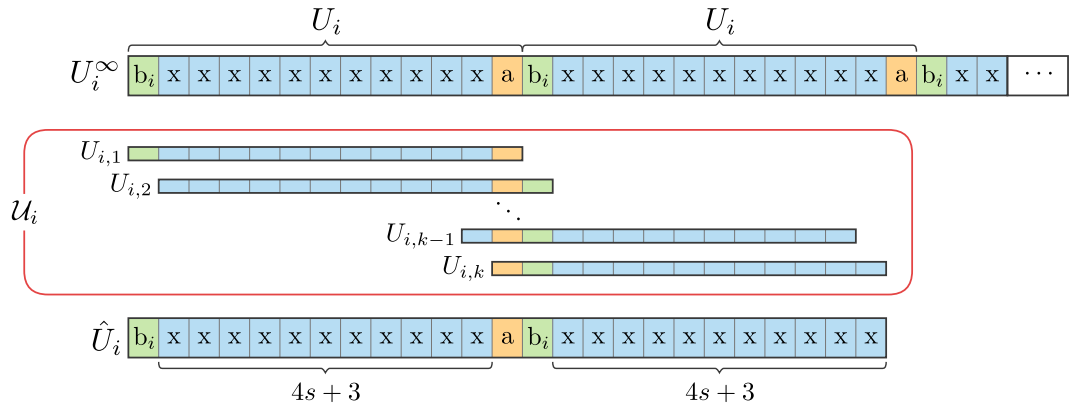
GREEDY が **悪い結合順序**を選んだときに、得られる解の文字列長が大きくなることを示す

具体的な流れ:

1. 各集合 $\mathcal{U}_i, \mathcal{V}_i$ に含まれる文字列をそれぞれ**悪い順序で**結合し、$2t$ 個の文字列 $\hat{U}_i, \hat{V}_i$ を得る
   - 各結合における overlap 長は $k-1$（＝ありえる最大値）
2. このようにして得られた文字列は**長い overlap を持たない**ため、最終的な解が長くなる

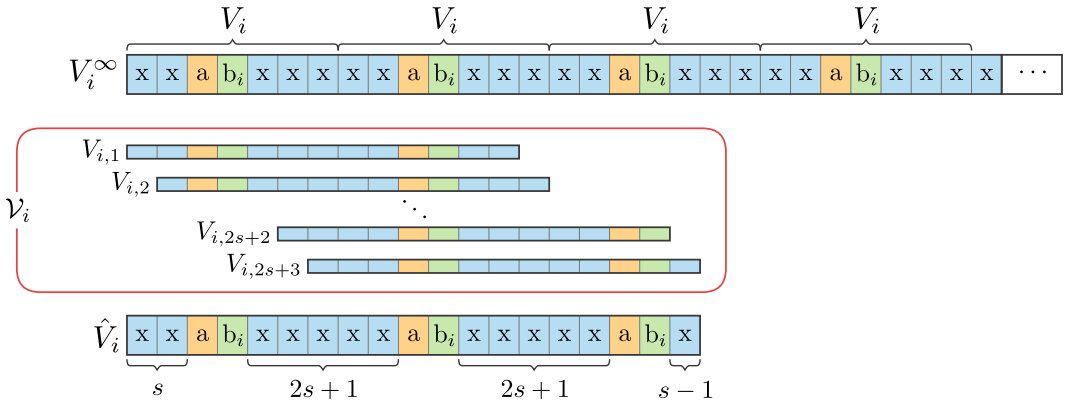
同グループに含まれる文字列の結合

1. $U_{i,1}, U_{i,2}, \ldots, U_{i,k}$ をこの順に結合し、$\hat{U}_i$ を得る



同グループに含まれる文字列の結合

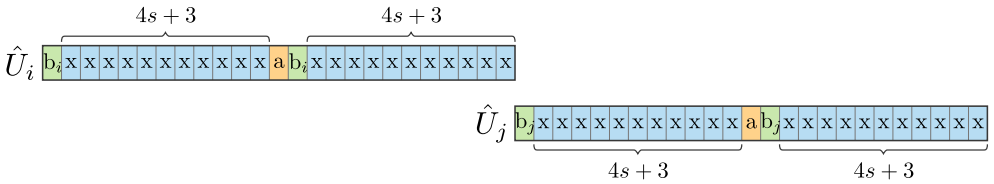
2. $V_{i,1}, V_{i,2}, \ldots, V_{i,2s+3}$ をこの順に結合し、$\hat{V}_i$ を得る



$\hat{U}_i, \hat{V}_i$ の overlap

$\hat{U}_\circ, \hat{V}_\circ$ の overlap 長の表は以下のようになる

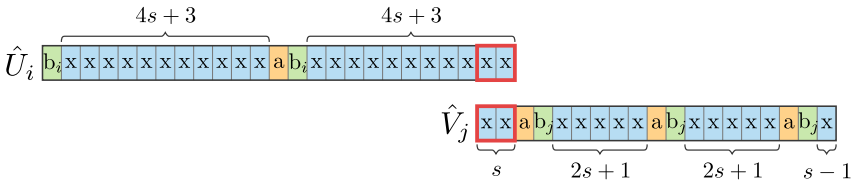
	$\hat{U}_j$	$\hat{V}_j$
$\hat{U}_i$	$0\ (i \neq j)$	s
$\hat{V}_i$	$\begin{cases} s & (i = j) \\ 0 & (i \neq j) \end{cases}$	$s-1\ (i \neq j)$



$\hat{U}_i, \hat{V}_i$ の overlap

$\hat{U}_\circ, \hat{V}_\circ$ の overlap 長の表は以下のようになる

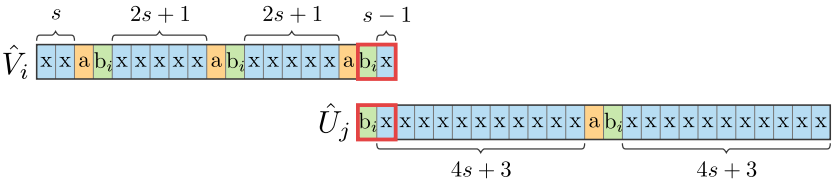
	$\hat{U}_j$	$\hat{V}_j$
$\hat{U}_i$	$0\ (i \neq j)$	s
$\hat{V}_i$	$\begin{cases} s & (i = j) \\ 0 & (i \neq j) \end{cases}$	$s-1\ (i \neq j)$



$\hat{U}_i, \hat{V}_i$ の overlap

$\hat{U}_\circ, \hat{V}_\circ$ の overlap 長の表は以下のようになる

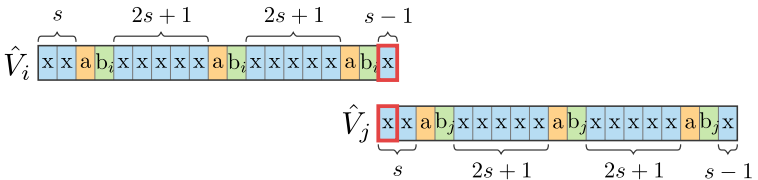
	$\hat{U}_j$	$\hat{V}_j$
$\hat{U}_i$	$0\ (i \neq j)$	s
$\hat{V}_i$	$\begin{cases} s & (i = j) \\ 0 & (i \neq j) \end{cases}$	$s-1\ (i \neq j)$



$\hat{U}_i, \hat{V}_i$ の overlap

$\hat{U}_\circ, \hat{V}_\circ$ の overlap 長の表は以下のようになる

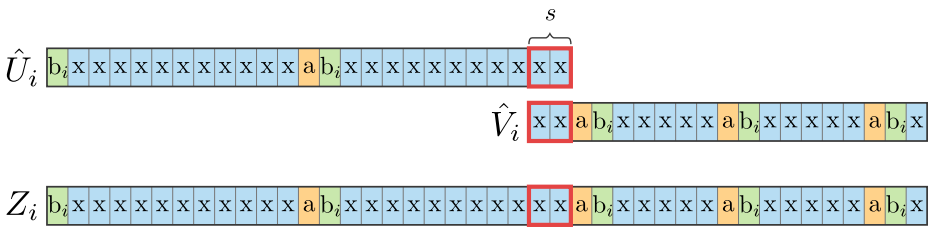
	$\hat{U}_j$	$\hat{V}_j$
$\hat{U}_i$	$0\ (i \neq j)$	s
$\hat{V}_i$	$\begin{cases} s & (i = j) \\ 0 & (i \neq j) \end{cases}$	$s-1\ (i \neq j)$



下界を与える GREEDY の実行例

以下のような順序で結合をしていくことで、最終的な出力文字列長を大きくできる

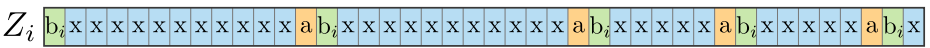
1. 各 $1 \leq i \leq t$ について、$\hat{U}_i, \hat{V}_i$ を結合して文字列 Z_i を得る（overlap 長 s ）
   - 最大の overlap をとっているため、GREEDY の条件を満たす
2. 各 Z_i を結合して最終的な文字列を得る（overlap 長 0 ）
   - 各 Z_j の先頭が Z_i に現れない文字 b_j なため、Z_i と Z_j は overlap を持たず、この結合も GREEDY の条件を満たす

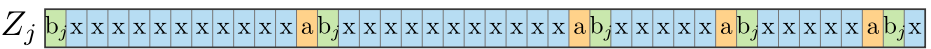


下界を与える GREEDY の実行例

以下のような順序で結合をしていくことで、最終的な出力文字列長を大きくできる

1. 各 $1 \leq i \leq t$ について、$\hat{U}_i, \hat{V}_i$ を結合して文字列 Z_i を得る（overlap 長 s ）
   - 最大の overlap をとっているため、GREEDY の条件を満たす
2. 各 Z_i を結合して最終的な文字列を得る（overlap 長 0 ）
   - 各 Z_j の先頭が Z_i に現れない文字 b_j なため、Z_i と Z_j は overlap を持たず、この結合も GREEDY の条件を満たす





GREEDY の出力文字列長の下界

最終的な文字列長の評価:

1. 長さ $k = 4s + 5$ の文字列を overlap 長 $k - 1$ でそれぞれ $4s + 5$ 個、$2s + 3$ 個結合するため、$|\hat{U}_i| = 8s + 9, |\hat{V}_i| = 6s + 7$
2. これらを overlap 長 s で結合するため、$|Z_i| = |\hat{U}_i| + |\hat{V}_i| - s = 13s + 16$
3. $Z_i, Z_j\ (i \neq j)$ は overlap を持たないため、最終的な文字列長は $(13s + 16)t$

定理 2: GREEDY の出力文字列長の下界

$\mathcal{S}$ に対する GREEDY のある実行では、長さ $(13s + 16)t$ の Common Superstring が得られる。

GREEDY の近似比下界

定理 1: Shortest Common Superstring 長の上界（再掲）

$\mathcal{S}$ は長さ $|\mathcal{S}| + k - 1$ の Common Superstring を持つ。

定理 2: GREEDY の出力文字列長の下界（再掲）

$\mathcal{S}$ に対する GREEDY のある実行では、長さ $(13s+16)t$ の Common Superstring が得られる。

$|\mathcal{S}| = (k+2s+3)t = (6s+8)t$ より、近似比の下界は

$$\frac{(13s+16)t}{(6s+8)t+4s+4} \xrightarrow{t\to\infty} \frac{13s+16}{6s+8} \xrightarrow{s\to\infty} \frac{13}{6}$$

定理 3: GREEDY の近似比下界（緩いバージョン）

GREEDY の近似比は $13/6$ 以上である。

目次

1. 近似比下界 $13/6 \geq 2.16$

2. 近似比下界 $9/4 = 2.25$

3. 近似比下界 $12/5 = 2.4$（概略）

今回の発表内容

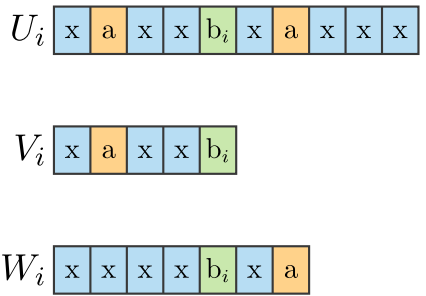
3 ステップ構成で紹介:

1. 近似比下界 $13/6 \geq 2.16$: 論文の結果を Simplify したもの
2. **近似比下界 $9/4 = 2.25$: 初期論文の結果**
   - **主な改善点: cycle を 2 つ→ 3 つに更新**
3. 近似比下界 $12/5 = 2.4$: 最新論文の結果（概略のみ）
   - 主な改善点: 巡回文字列構成からの脱却

文字列集合 S の構成 (1/3)

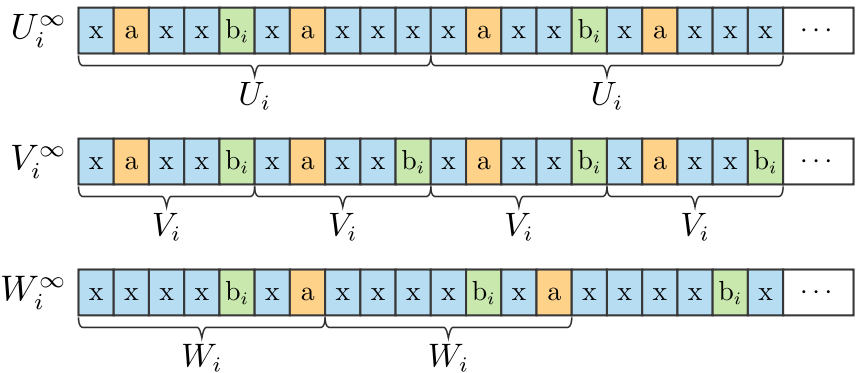
パラメータ $k = 2s \geq 10$ と $t \geq 1$ をもとに、$3t$ 個の文字列 U_i, V_i, W_i $(1 \leq i \leq t)$ を以下のように定義

	長さ	a の位置	b_i の位置
U_i	$2s$	$2,\ s+2$	s
V_i	s	2	s
W_i	$s+2$	$s+2$	s



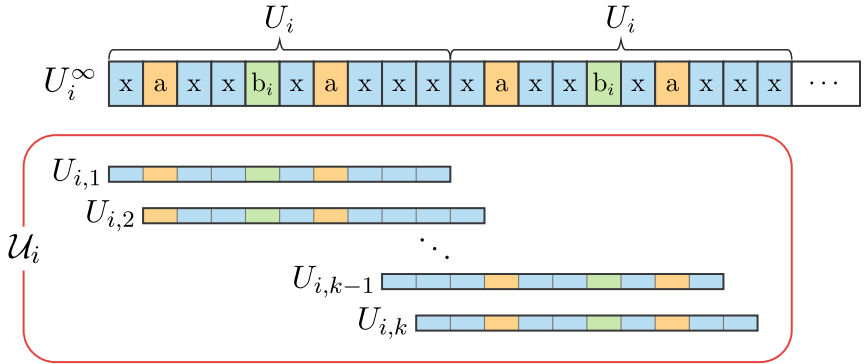
文字列集合 $\mathcal{S}$ の構成 (2/3)

構成した文字列から、巡回文字列 $U_i^\infty, V_i^\infty, W_i^\infty$ を得る



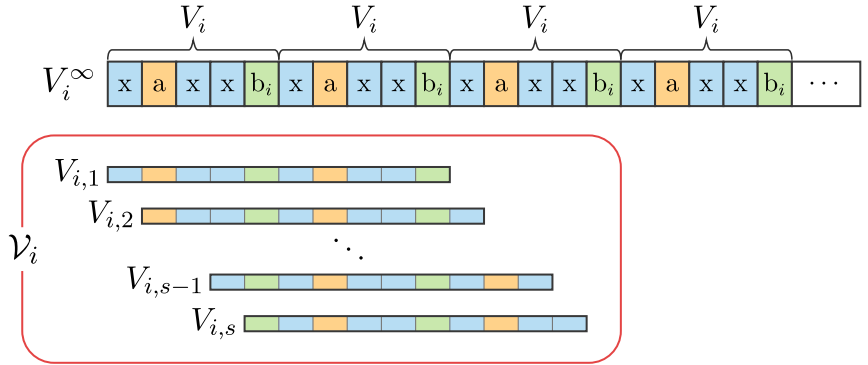
文字列集合 $\mathcal{S}$ の構成 (3/3)

各巡回文字列から長さ k の部分文字列を切り出し、これらを入力文字列集合 $\mathcal{S} = (\mathcal{U}_1 \sqcup \cdots \sqcup \mathcal{U}_t) \sqcup (\mathcal{V}_1 \sqcup \cdots \sqcup \mathcal{V}_t) \sqcup (\mathcal{W}_1 \sqcup \cdots \sqcup \mathcal{W}_t)$ とする



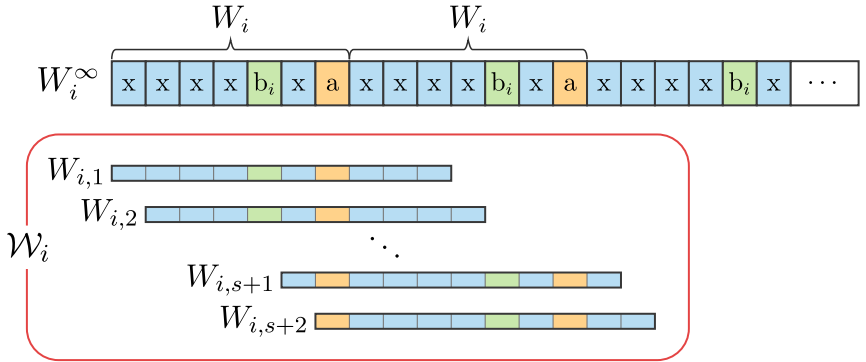
文字列集合 $\mathcal{S}$ の構成 (3/3)

各巡回文字列から長さ k の部分文字列を切り出し、これらを入力文字列集合 $\mathcal{S} = (\mathcal{U}_1 \sqcup \cdots \sqcup \mathcal{U}_t) \sqcup (\mathcal{V}_1 \sqcup \cdots \sqcup \mathcal{V}_t) \sqcup (\mathcal{W}_1 \sqcup \cdots \sqcup \mathcal{W}_t)$ とする



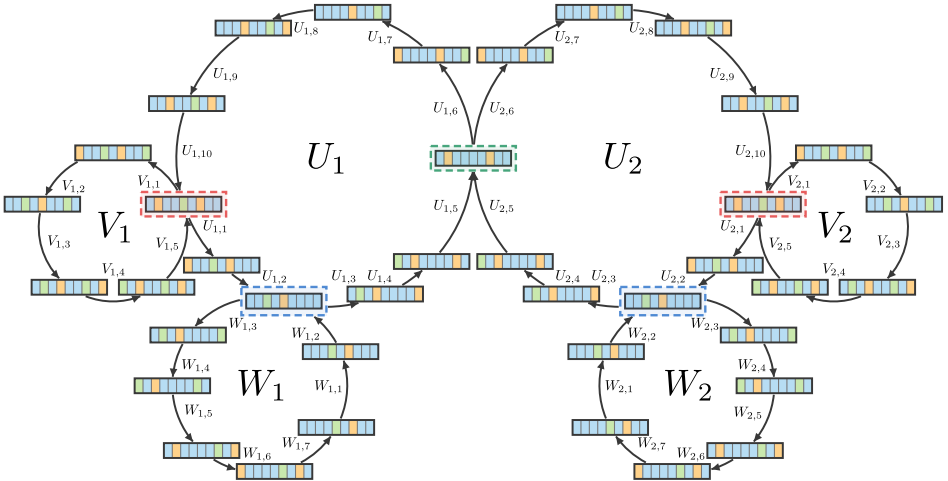
文字列集合 $\mathcal{S}$ の構成 (3/3)

各巡回文字列から長さ k の部分文字列を切り出し、これらを入力文字列集合 $\mathcal{S} = (\mathcal{U}_1 \sqcup \cdots \sqcup \mathcal{U}_t) \sqcup (\mathcal{V}_1 \sqcup \cdots \sqcup \mathcal{V}_t) \sqcup (\mathcal{W}_1 \sqcup \cdots \sqcup \mathcal{W}_t)$ とする



文字列集合 $\mathcal{S}$ の Rauzy Graph

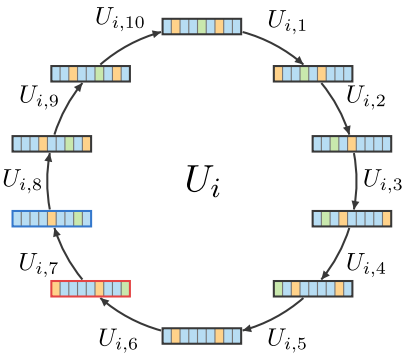
事実: $\mathcal{S}$ の Rauzy Graph は以下のようになり、オイラー路を持つ

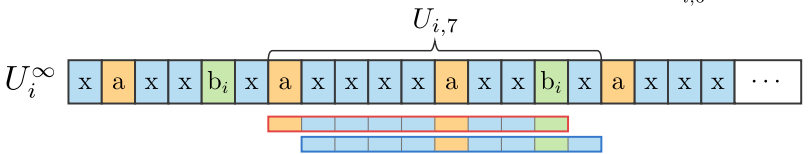


Rauzy Graphにおけるサイクル

各文字列 $U_{i,j}$ は U_i^∞ の長さ k の部分文字列なため、各辺集合 $\mathcal{U}_i$ はサイクルを成す

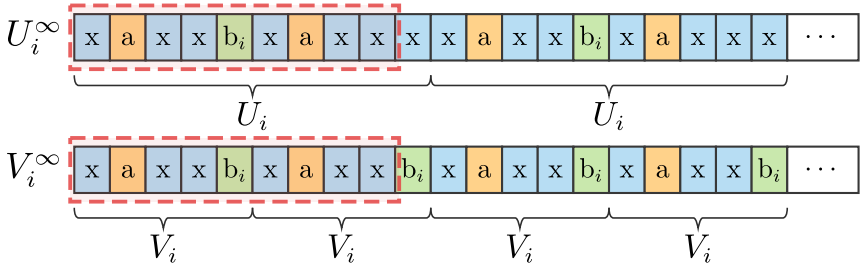
- V_i, W_i に関する部分も同様





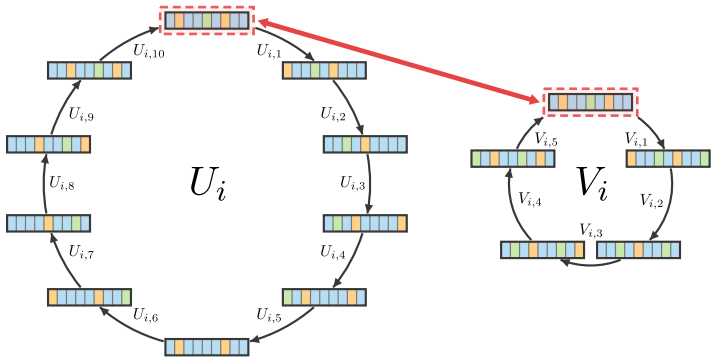
共通頂点によるサイクルの結合 (1/3)

(U_i^∞, V_i^∞) は長さ $k-1$ の共通部分文字列を持つ
→ U_i と V_i のサイクルが結合できる



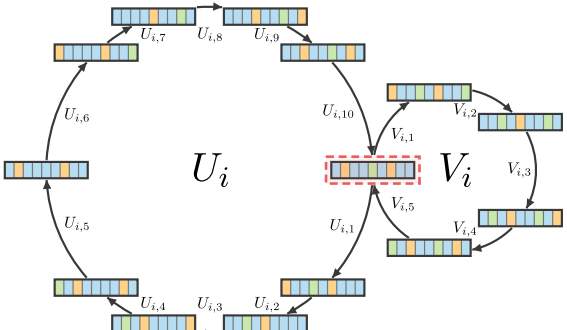
共通頂点によるサイクルの結合 (1/3)

$(U_i^{\infty}, V_i^{\infty})$ は長さ $k-1$ の共通部分文字列を持つ
→ U_i と V_i のサイクルが結合できる



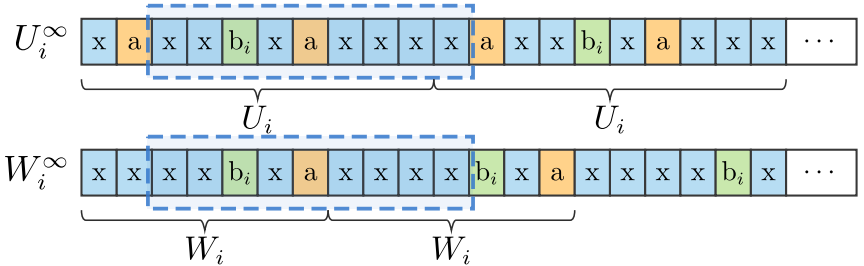
共通頂点によるサイクルの結合 (1/3)

$(U_i^{\infty}, V_i^{\infty})$ は長さ $k-1$ の共通部分文字列を持つ
→ U_i と V_i のサイクルが結合できる



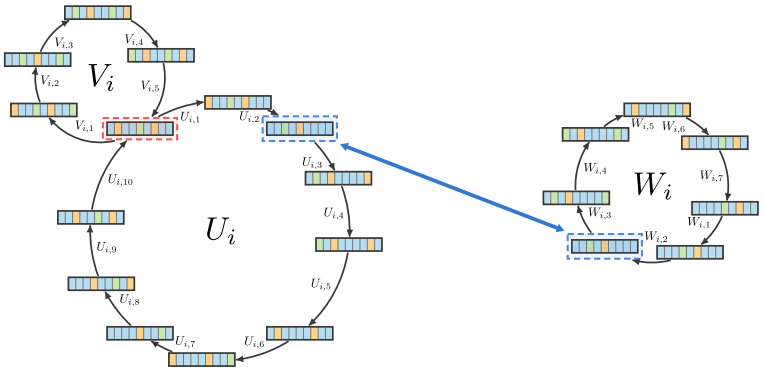
共通頂点によるサイクルの結合 (2/3)

(U_i^∞, W_i^∞) は長さ $k-1$ の共通部分文字列を持つ
→ U_i, W_i のサイクルが結合できる



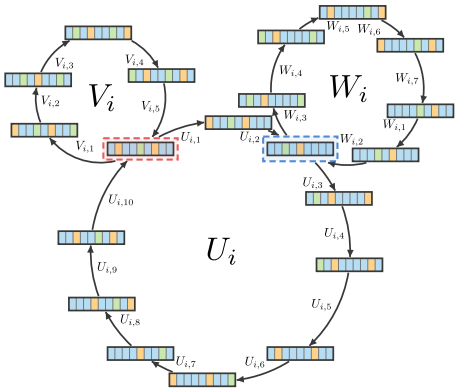
共通頂点によるサイクルの結合 (2/3)

$(U_i^{\infty}, W_i^{\infty})$ は長さ $k-1$ の共通部分文字列を持つ
→ U_i, W_i のサイクルが結合できる



共通頂点によるサイクルの結合 (2/3)

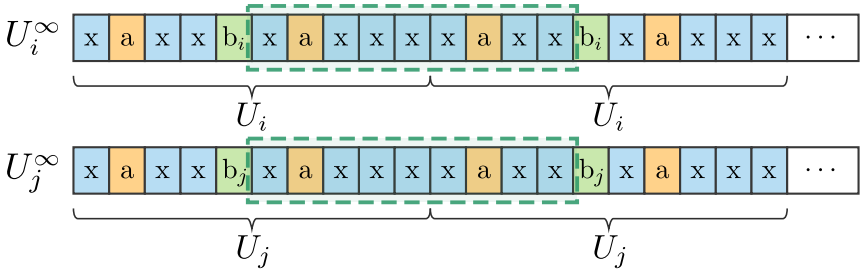
$(U_i^{\infty}, W_i^{\infty})$ は長さ $k-1$ の共通部分文字列を持つ
→ U_i, W_i のサイクルが結合できる



共通頂点によるサイクルの結合 (3/3)

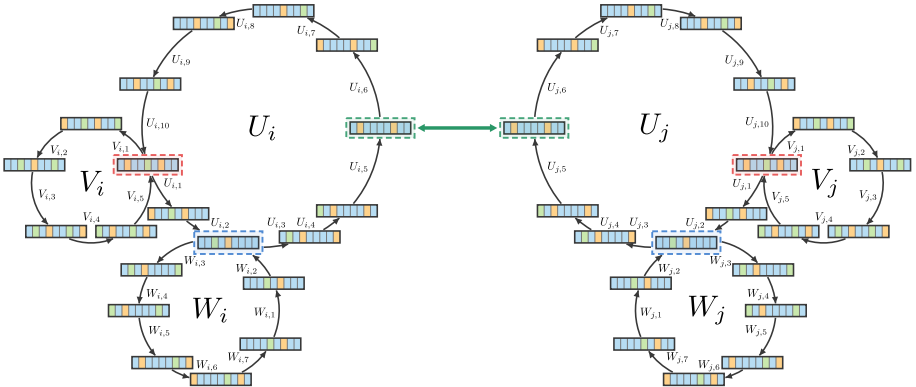
(U_i^∞, U_j^∞) は長さ $k-1$ の共通部分文字列を持つ

→ i, j それぞれのグラフが結合できる



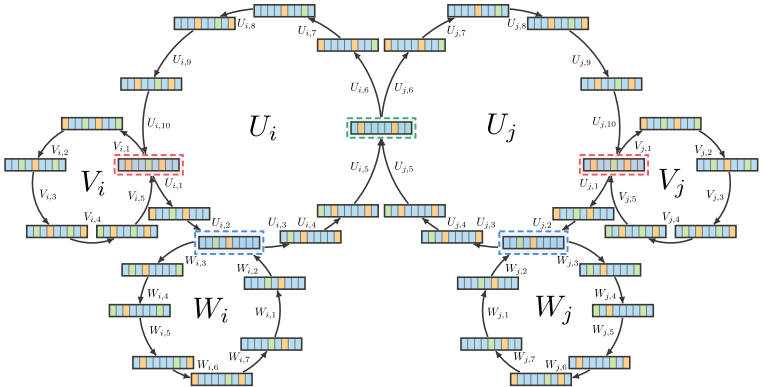
共通頂点によるサイクルの結合 (3/3)

$(U_i^{\infty}, U_j^{\infty})$ は長さ $k-1$ の共通部分文字列を持つ
→ i, j それぞれのグラフが結合できる



共通頂点によるサイクルの結合 (3/3)

$(U_i^{\infty}, U_j^{\infty})$ は長さ $k-1$ の共通部分文字列を持つ
→ i, j それぞれのグラフが結合できる

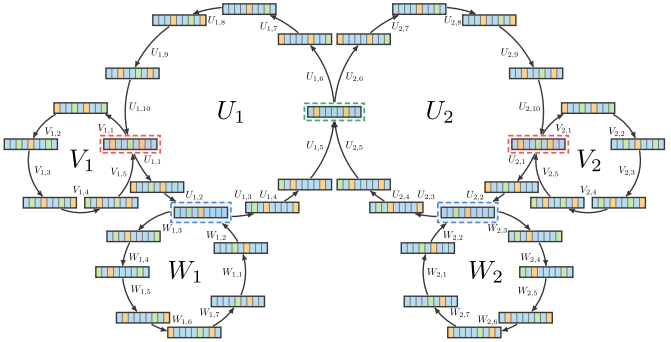


SCS 長の上界

$\mathcal{S}$ の Rauzy Graph がオイラー路を持つため、以下が成り立つ

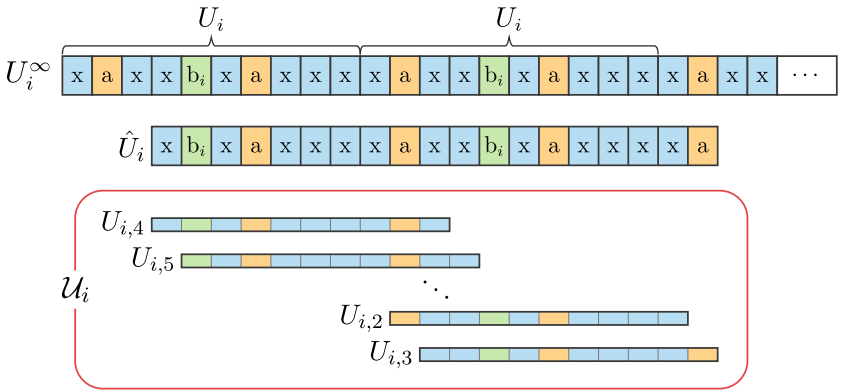
定理 4: Shortest Common Superstring 長の上界

$\mathcal{S}$ は長さ $|\mathcal{S}| + k - 1$ の Common Superstring を持つ。



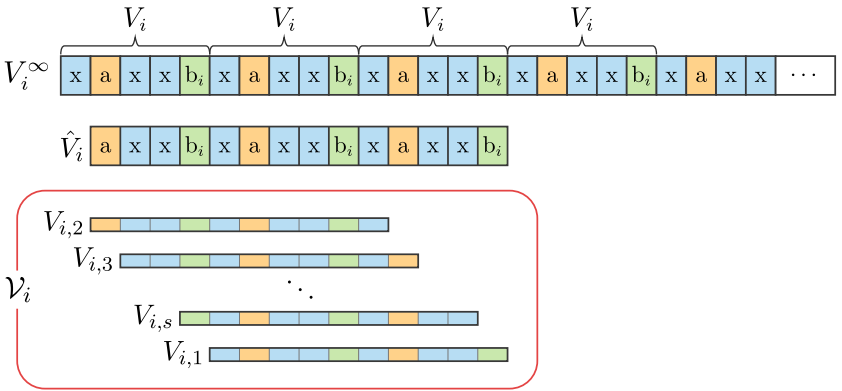
同グループに含まれる文字列の結合

1. $U_{i,4}, U_{i,5}, \ldots, U_{i,3}$ をこの順に結合し、$\hat{U}_i$ を得る



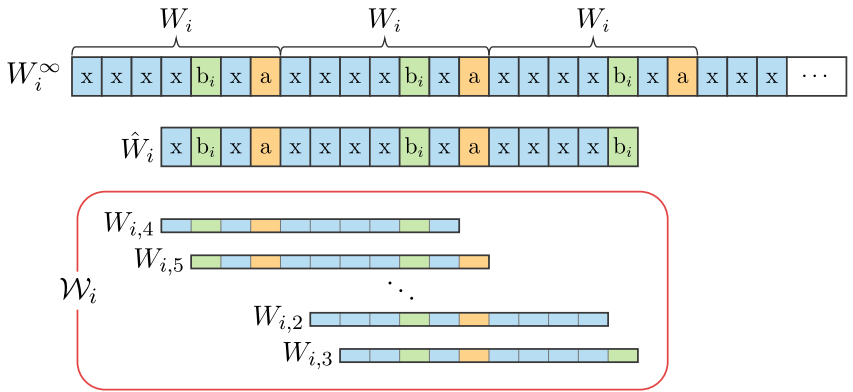
同グループに含まれる文字列の結合

2. $V_{i,2}, V_{i,3}, \ldots, V_{i,1}$ をこの順に結合し、$\hat{V}_i$ を得る



同グループに含まれる文字列の結合

3. $W_{i,4}, W_{i,5}, \ldots, W_{i,3}$ をこの順に結合し、$\hat{W}_i$ を得る

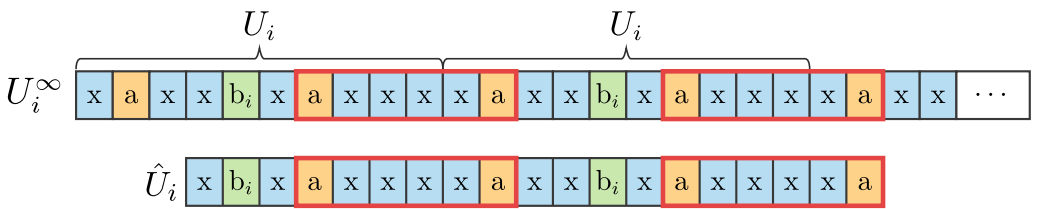


結合順序の選び方に関する直感

$U_i^\infty, V_i^\infty, W_i^\infty$ が持つ**特徴的な部分文字列**に着目する

- U_i^∞**:** $\mathrm{a}\mathrm{x}^{s-1}\mathrm{a}$ **は** $U_\circ^\infty$ **にしか現れない**
- V_i^∞: $\mathrm{a}\mathrm{x}^{s-3}\mathrm{b}_i$ は U_i^∞, V_i^∞ にしか現れない
- W_i^∞: $\mathrm{a}\mathrm{x}^{s-1}\mathrm{b}_i$ は W_i^∞ にしか現れない

$\hat{U}_i, \hat{V}_i, \hat{W}_i$ は、これらが文字列末尾に来るよう恣意的に構成している

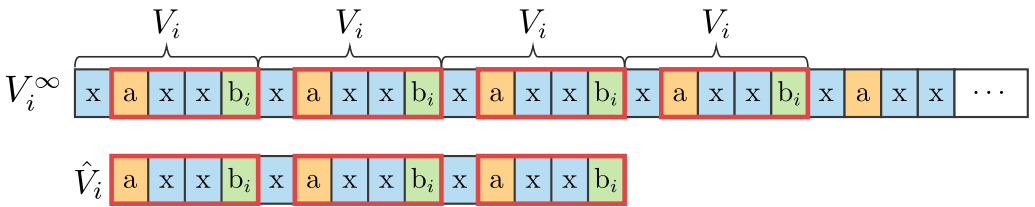


結合順序の選び方に関する直感

$U_i^\infty, V_i^\infty, W_i^\infty$ が持つ**特徴的な部分文字列**に着目する

- U_i^∞: $\mathrm{a}\mathrm{x}^{s-1}\mathrm{a}$ は $U_\circ^\infty$ にしか現れない
- **V_i^∞: $\mathrm{a}\mathrm{x}^{s-3}\mathrm{b}_i$ は U_i^∞, V_i^∞ にしか現れない**
- W_i^∞: $\mathrm{a}\mathrm{x}^{s-1}\mathrm{b}_i$ は W_i^∞ にしか現れない

$\hat{U}_i, \hat{V}_i, \hat{W}_i$ は、これらが文字列末尾に来るよう恣意的に構成している

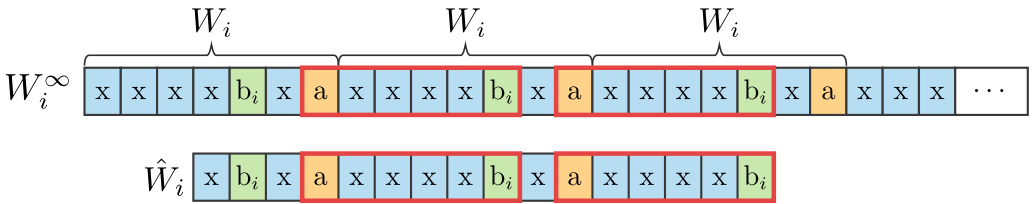


結合順序の選び方に関する直感

$U_i^\infty, V_i^\infty, W_i^\infty$ が持つ**特徴的な部分文字列**に着目する

- U_i^∞: $\mathrm{a}\mathrm{x}^{s-1}\mathrm{a}$ は $U_\circ^\infty$ にしか現れない
- V_i^∞: $\mathrm{a}\mathrm{x}^{s-3}\mathrm{b}_i$ は U_i^∞, V_i^∞ にしか現れない
- W_i^∞: $\mathrm{a}\mathrm{x}^{s-1}\mathrm{b}_i$ は W_i^∞ にしか現れない

$\hat{U}_i, \hat{V}_i, \hat{W}_i$ は、これらが文字列末尾に来るよう恣意的に構成している

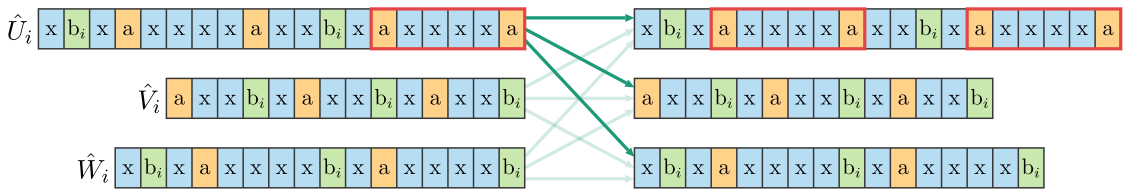


$\hat{U}_i, \hat{V}_i, \hat{W}_i$ の overlap (1/3)

$\hat{U}_i, \hat{V}_i, \hat{W}_i$ 同士に長い overlap がないことの確認:

1. $\hat{U}_i$ を末尾にする場合

- $\hat{U}_i$ は $\mathrm{a}\mathrm{x}^{s-1}\mathrm{a}$ で終わるが、この部分を共有できる候補は（自身を除き）$\hat{U}_j$ $(i \neq j)$ のみ
- b_i の存在により、$\hat{U}_j$ とも長い overlap を持てない

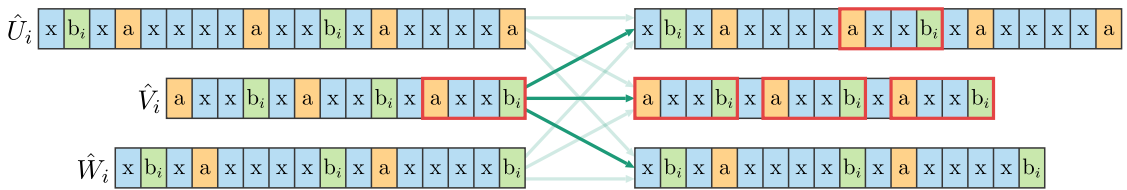


$\hat{U}_i, \hat{V}_i, \hat{W}_i$ の overlap (1/3)

$\hat{U}_i, \hat{V}_i, \hat{W}_i$ 同士に長い overlap がないことの確認:

2. $\hat{V}_i$ を末尾にする場合

- $\hat{V}_i$ は $\mathrm{a}\mathrm{x}^{s-3}\mathrm{b}_i$ で終わるが、この部分を共有できる候補は（自身を除き）$\hat{U}_i$ のみ
- $\hat{U}_i$ 中の $\mathrm{a}\mathrm{x}^{s-1}\mathrm{a}$ の存在より、$\hat{U}_i$ とも長い overlap を持てない

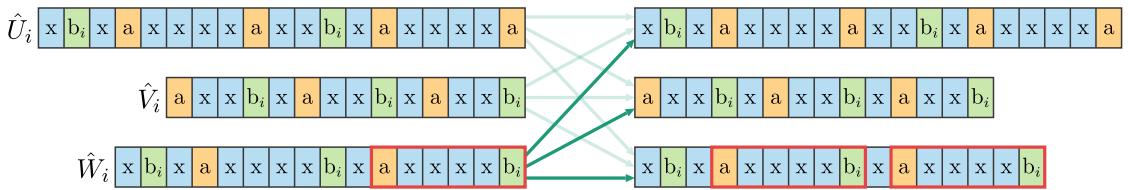


$\hat{U}_i, \hat{V}_i, \hat{W}_i$ の overlap (1/3)

$\hat{U}_i, \hat{V}_i, \hat{W}_i$ 同士に長い overlap がないことの確認:

3. $\hat{W}_i$ を末尾にする場合

- $\hat{W}_i$ は $\mathrm{a}\mathrm{x}^{s-1}\mathrm{b}_i$ で終わるが、この部分は他の文字列に現れないため、長い overlap を持てない

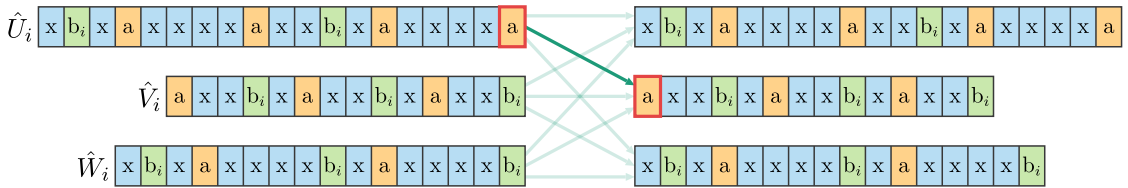


$\hat{U}_i, \hat{V}_i, \hat{W}_i$ の overlap (2/3)

$\hat{U}_i, \hat{V}_i, \hat{W}_i$ 同士の短い overlap の列挙:

A. $(\hat{U}_\circ, \hat{V}_\circ)$ の組が 1 文字分 (a) の overlap を持つ

B. $(\hat{V}_i, \hat{U}_i),\ (\hat{V}_i, \hat{W}_i),\ (\hat{W}_i, \hat{U}_i)$ の組が $s-3$ 文字分 $(\mathrm{x}^{s-4}\mathrm{b}_i)$ の overlap を持つ

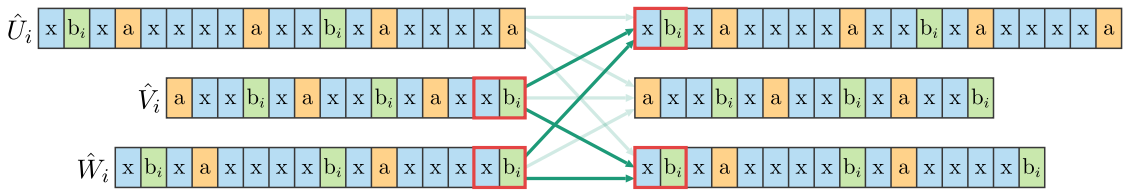


$\hat{U}_i, \hat{V}_i, \hat{W}_i$ の overlap (2/3)

$\hat{U}_i, \hat{V}_i, \hat{W}_i$ 同士の短い overlap の列挙:

A. $(\hat{U}_\circ, \hat{V}_\circ)$ の組が 1 文字分 (a) の overlap を持つ

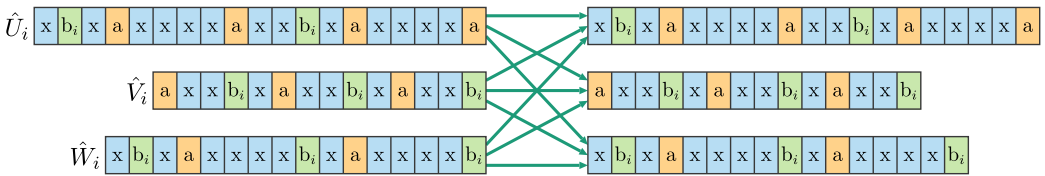
B. $(\hat{V}_i, \hat{U}_i),\ (\hat{V}_i, \hat{W}_i),\ (\hat{W}_i, \hat{U}_i)$ の組が $s-3$ 文字分 $(\mathrm{x}^{s-4}\mathrm{b}_i)$ の overlap を持つ



$\hat{U}_i, \hat{V}_i, \hat{W}_i$ の overlap (3/3)

$\hat{U}_\circ, \hat{V}_\circ, \hat{W}_\circ$ の overlap 長の表

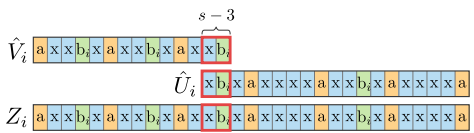
	$\hat{U}_j$	$\hat{V}_j$	$\hat{W}_j$
$\hat{U}_i$	$0\ (i \neq j)$	1	0
$\hat{V}_i$	$\begin{cases} s-3 & (i=j) \\ 0 & (i \neq j) \end{cases}$	$0\ (i \neq j)$	$\begin{cases} s-3 & (i=j) \\ 0 & (i \neq j) \end{cases}$
$\hat{W}_i$	$\begin{cases} s-3 & (i=j) \\ 0 & (i \neq j) \end{cases}$	0	$0\ (i \neq j)$



下界を与える GREEDY の実行例

以下のような順序で結合をしていくことで、最終的な出力文字列長を大きくできる

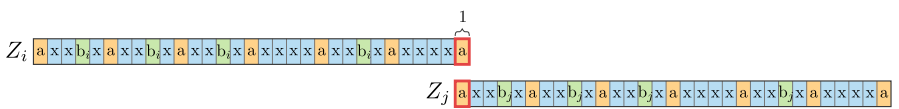
1. 各 $1 \leq i \leq t$ について、$\hat{V}_i, \hat{U}_i$ を結合して文字列 Z_i を得る（overlap 長 $s-3$）
   - 最大の overlap をとっているため、GREEDY の条件を満たす
2. 各 Z_i を結合して文字列 Z を得る（overlap 長 1）
   - このとき overlap 長の最大値は 1 なため、GREEDY の条件を満たす
3. Z と各 $\hat{W}_i$ を結合して最終的な文字列を得る（overlap 長 0）
   - $Z, \hat{W}_1, \ldots, \hat{W}_t$ は overlap を持たず、この結合も GREEDY の条件を満たす



下界を与える GREEDY の実行例

以下のような順序で結合をしていくことで、最終的な出力文字列長を大きくできる

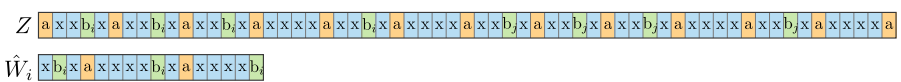
1. 各 $1 \leq i \leq t$ について、$\hat{V}_i, \hat{U}_i$ を結合して文字列 Z_i を得る（overlap 長 $s-3$）
   - 最大の overlap をとっているため、GREEDY の条件を満たす
2. **各 Z_i を結合して文字列 Z を得る（overlap 長 1）**
   - **このとき overlap 長の最大値は 1 なため、GREEDY の条件を満たす**
3. Z と各 $\hat{W}_i$ を結合して最終的な文字列を得る（overlap 長 0）
   - $Z, \hat{W}_1, \ldots, \hat{W}_t$ は overlap を持たず、この結合も GREEDY の条件を満たす



下界を与える GREEDY の実行例

以下のような順序で結合をしていくことで、最終的な出力文字列長を大きくできる

1. 各 $1 \leq i \leq t$ について、$\hat{V}_i, \hat{U}_i$ を結合して文字列 Z_i を得る（overlap 長 $s-3$）
   - 最大の overlap をとっているため、GREEDY の条件を満たす
2. 各 Z_i を結合して文字列 Z を得る（overlap 長 1）
   - このとき overlap 長の最大値は 1 なため、GREEDY の条件を満たす
3. **Z と各 $\hat{W}_i$ を結合して最終的な文字列を得る（overlap 長 0）**
   - **$Z, \hat{W}_1, \ldots, \hat{W}_t$ は overlap を持たず、この結合も GREEDY の条件を満たす**



GREEDYの出力文字列長の下界

最終的な文字列長の評価:

1. 長さ $k = 2s$ の文字列を overlap 長 $k - 1$ でそれぞれ $2s$ 個、s 個、$s + 2$ 個結合するため、$|\hat{U}_i| = 4s - 1$, $|\hat{V}_i| = 3s - 1$, $|\hat{W}_i| = 3s + 1$
2. これらを overlap 長 $s - 3$ で結合するため、$|Z_i| = |\hat{V}_i| + |\hat{U}_i| - (s - 3) = 6s + 1$
3. 各 Z_i を overlap 長 1 で結合するため、$|Z| = (6s + 1)t - (t - 1) = 6st + 1$
4. Z と各 $\hat{W}_i$ は overlap を持たないため、最終的な文字列長は $(6st + 1) + (3s + 1)t = (9s + 1)t + 1$

定理 5: GREEDYの出力文字列長の下界

$\mathcal{S}$ に対する GREEDY のある実行では、長さ $(9s + 1)t + 1$ の Common Superstring が得られる。

GREEDY の近似比下界

定理 4: Shortest Common Superstring 長の上界（再掲）

$\mathcal{S}$ は長さ $|\mathcal{S}| + k - 1$ の Common Superstring を持つ。

定理 5: GREEDY の出力文字列長の下界（再掲）

$\mathcal{S}$ に対する GREEDY のある実行では、長さ $(9s+1)t+1$ の Common Superstring が得られる。

$|\mathcal{S}| = (k+s+s+2)t = (4s+2)t$ より、近似比の下界は

$$\frac{(9s+1)t+1}{(4s+2)t+2s-1} \xrightarrow{t\to\infty} \frac{9s+1}{4s+2} \xrightarrow{s\to\infty} \frac{9}{4}$$

定理 6: GREEDY の近似比下界

GREEDY の近似比は $9/4$ 以上である。

目次

1. 近似比下界 $13/6 \geq 2.16$

2. 近似比下界 $9/4 = 2.25$

3. 近似比下界 $12/5 = 2.4$（概略）

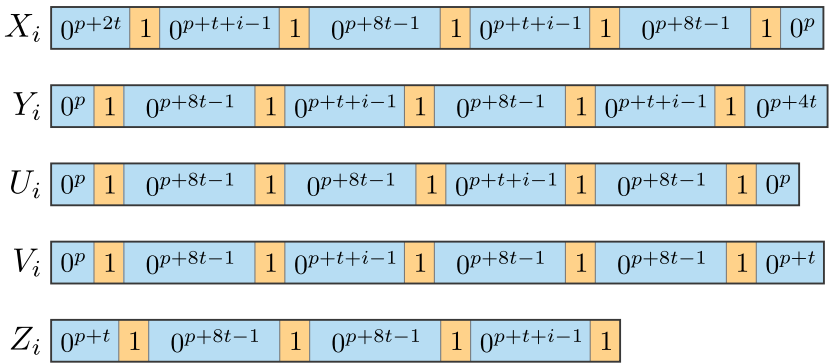
今回の発表内容

3 ステップ構成で紹介:

1. 近似比下界 $13/6 \geq 2.16$: 論文の結果を Simplify したもの
2. 近似比下界 $9/4 = 2.25$: 初期論文の結果
   - 主な改善点: cycle を 2 つ→ 3 つに更新
3. **近似比下界 $12/5 = 2.4$: 最新論文の結果（概略のみ）**
   - **主な改善点: 巡回文字列構成からの脱却**

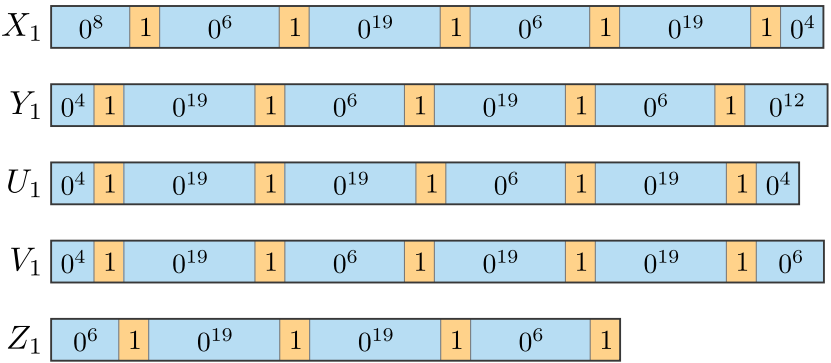
文字列集合 $\mathcal{S}$ の構成

パラメータ $t \geq 1, p = t^2$ をもとに、$5t$ 個の文字列 $X_i, Y_i, U_i, V_i, Z_i\ (1 \leq i \leq t)$ を構成



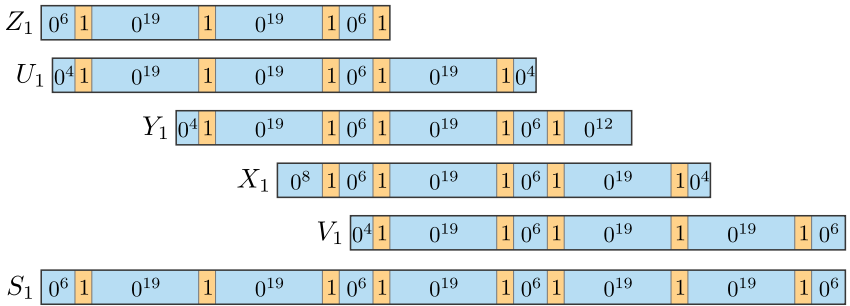
文字列集合 $\mathcal{S}$ の構成

パラメータ $t \geq 1, p = t^2$ をもとに、$5t$ 個の文字列 $X_i, Y_i, U_i, V_i, Z_i\ (1 \leq i \leq t)$ を構成



「良い順序」での結合

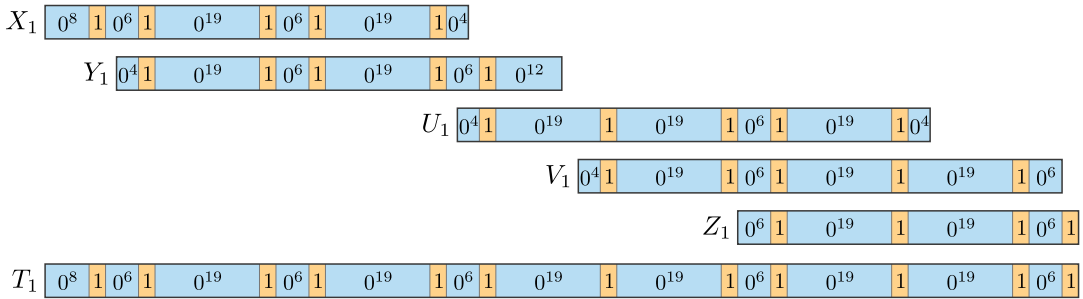
良い実行では、Z_i, U_i, Y_i, X_i, V_i の順で結合することで文字列 S_i が得られる



得られる文字列の長さは $|S_i| = 9p + O(t)$

「悪い順序」での結合

GREEDY の特定の実行では、X_i, Y_i, U_i, V_i, Z_i の順で結合され、T_i が得られる



得られる文字列の長さは $|T_i| = 12p + O(t)$

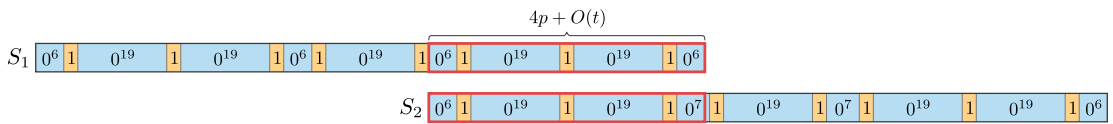
S_i の結合

良い実行の場合:

- S_i の長さ: $9p + O(t)$
- S_i, S_j の overlap 長: $4p + O(t)$

t 個の結合後の合計長は $5pt + O(t^2)$

（長さ $9p$ のうち $4p$ 部分が重複するため）

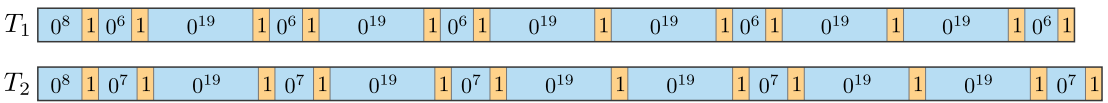


T_i の結合

悪い実行の場合:

- T_i の長さ: $12p + O(t)$
- T_i, T_j の overlap 長: 0

overlap 長が 0 なため、t 個の結合後の合計長は $12pt + O(t^2)$



2 つの実行の長さ比

ここまでの長さ議論より、2 つの実行の長さ比は

$$\frac{12pt + O(t^2)}{5pt + O(p + t^2)} = \frac{12t^3 + O(t^2)}{5t^3 + O(t^2)} \xrightarrow{t\to\infty} \frac{12}{5}$$

したがって、以下の定理が成り立つ:

定理 7: GREEDY の近似比下界

GREEDY の近似比は $12/5$ 以上である。