# Partial Word Representation

F. Blanchet-Sadri

Fields Institute Workshop

This material is based upon work supported by the National Science Foundation under Grant No. DMS–1060775.

# Partial words and compatibility

- A partial word is a sequence that may have undefined positions, called holes and denoted by $\diamond$'s, that match any letter of the alphabet $A$ over which it is defined (a full word is a partial word without holes); we also say that $\diamond$ is compatible with each $a \in A$.

  $a\diamond b\diamond aab$ is a partial word with two holes over $\{a, b\}$

- Two partial words $w$ and $w'$ of equal length are compatible, denoted by $w \uparrow w'$, if $w[i] = w'[i]$ whenever $w[i], w'[i] \in A$.

$$\begin{array}{cccccc} & a & \diamond & b & \diamond & a \\ \uparrow & & & & & \\ & \diamond & \diamond & b & a & a \end{array} \qquad \begin{array}{cccccc} & a & \diamond & b & \diamond & a \\ \not\uparrow & & & & & \\ & \diamond & \diamond & a & a & a \end{array}$$

# Factor and subword

- A partial word $u$ is a factor of the partial word $w$ if $u$ is a block of consecutive symbols of $w$.

  $\diamond a \diamond$ is a factor of $aa\diamond a\diamond b$

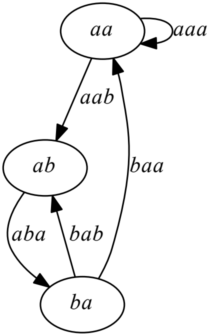
- A full word $u$ is a subword of the partial word $w$ if it is compatible with a factor of $w$.

  $aaa$, $aab$, $baa$, $bab$ are the subwords of $aa\diamond a\diamond b$ corresponding to the factor $\diamond a \diamond$

# Some computational problems

- We define REP, or the problem of deciding whether a set $S$ of words of length $n$ is representable, i.e., whether $S = \text{sub}_w(n)$ for some integer $n$ and partial word $w$.
- If $h$ is a non-negative integer, we also define $h$-REP, or the problem of deciding whether $S$ is $h$-representable, i.e., whether $S = \text{sub}_w(n)$ for some integer $n$ and partial word $w$ with exactly $h$ holes.

# Rauzy graph of $S = \{aaa, aab, aba, baa, bab\}$




$S$ is 0-representable by $w = aaababaa$

# Why partial words? (Compression of representations)

$$S = \{aaa, aab, aba, baa, bab\}$$

is representable by

*aaababaa*

# Why partial words? (Compression of representations)

$S = \{aaa, aab, aba, baa, bab\}$

is representable by

$\diamond aabab$

# Why partial words? (Compression of representations)

$$S = \{aaa, aab, aba, baa, bab\}$$

is representable by

$$a\diamond a\diamond$$

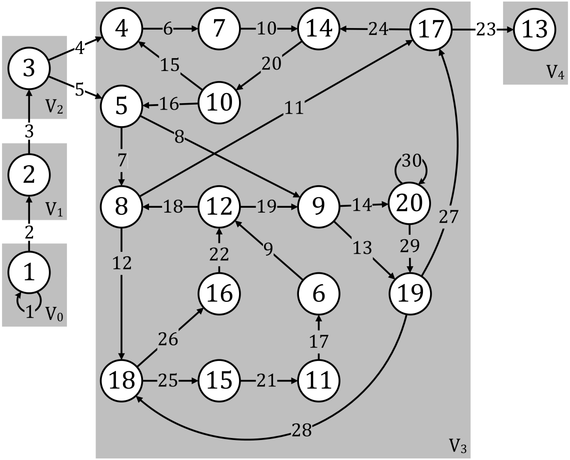
# Why partial words? (Representation of non-0-representable sets)

- Set $S$ of 30 words of length six:

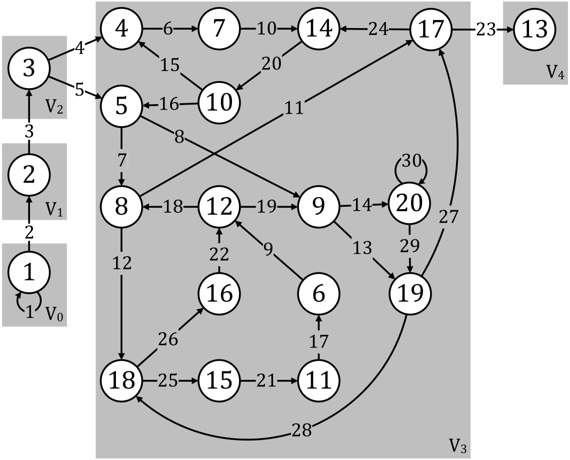
| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *aaaaaa* | 6 | *aabbaa* | 11 | *abbbaa* | 16 | *baabbb* | 21 | *bbabab* | 26 | *bbbabb* |
| 2 | *aaaaab* | 7 | *aabbba* | 12 | *abbbab* | 17 | *bababb* | 22 | *bbabbb* | 27 | *bbbbaa* |
| 3 | *aaaabb* | 8 | *aabbbb* | 13 | *abbbba* | 18 | *babbba* | 23 | *bbbaaa* | 28 | *bbbbab* |
| 4 | *aaabba* | 9 | *ababbb* | 14 | *abbbbb* | 19 | *babbbb* | 24 | *bbbaab* | 29 | *bbbbba* |
| 5 | *aaabbb* | 10 | *abbaab* | 15 | *baabba* | 20 | *bbaabb* | 25 | *bbbaba* | 30 | *bbbbbb* |

- Rauzy graph $(V, E)$ of $S$, where $E = S$ and $V = \text{sub}_S(5)$ consists of 20 words of length five:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | *aaaaa* | 5 | *aabbb* | 9 | *abbbb* | 13 | *bbaaa* | 17 | *bbbaa* |
| 2 | *aaaab* | 6 | *ababb* | 10 | *baabb* | 14 | *bbaab* | 18 | *bbbab* |
| 3 | *aaabb* | 7 | *abbaa* | 11 | *babab* | 15 | *bbaba* | 19 | *bbbba* |
| 4 | *aabba* | 8 | *abbba* | 12 | *babbb* | 16 | *bbabb* | 20 | *bbbbb* |




| | |
|---|---|
| 4 | *aaabba* |
| 5 | *aaabbb* |
| | *aaabb*◇ |




4 *aaabba*

5 *aaabbb*

---

*aaabb*◇

# Membership of $h$-Rep in $\mathcal{P}$

**Theorem**

*$h$-Rep is in $\mathcal{P}$ for any fixed non-negative integer $h$.*

---

F. Blanchet-Sadri and S. Simmons, Deciding representability of sets of words of equal length. *Theoretical Computer Science* **475** (2013) 34–46

# Membership of REP in $\mathcal{P}$

**Theorem**
REP *is in* $\mathcal{P}$.

F. Blanchet-Sadri and S. Munteanu, Deciding representability of sets of words of equal length in polynomial time. Submitted

# Open problems

- Characterize the sets of words that are representable.
- Characterize minimal representing partial words (can they be constructed efficiently?)