Factorisation forests for infinite words

Thomas Colcombet
Cnrs, Irisa
Fct 2007

OVERVIEW

- Factorisation forest theorems
  - Original presentation, by trees
  - By Ramseyan regular expressions
  - By splits
- An elementary application
- Ideas about the proof
- Extension to infinite words
- Application to automata over scattered linear orderings
- Related work

SEMIGROUPS

Semigroup: $(S, .)$, with multiplication associative.
E.g., $(A^*, .)$ is a semigroup.
Morphism: $f : S \to S'$ with $f(x.y) = f(x).f(y)$

SEMIGROUPS

Semigroup: $(S, .)$, with multiplication associative.
E.g., $(A^*, .)$ is a semigroup.
Morphism: $f : S \to S'$ with $f(x.y) = f(x).f(y)$

Th(Kleene): A language is regular iff it is of the form $f^{-1}(M)$ for $M \subseteq S$, with S a finite semigroup.

E.g., Words over $\{a, b\}$ containing an even number of a's is $f^{-1}(0)$ with
Semigroup: $S = \{0, 1\}$, $00 = 11 = 0$, $10 = 01 = 1$
Morphism: $f(a) = 1$, $f(b) = 0$

Semigroups

Semigroup: $(S, .)$, with multiplication associative.
E.g., $(A^*, .)$ is a semigroup.
Morphism: $f : S \rightarrow S'$ with $f(x.y) = f(x).f(y)$

Th(Kleene): A language is regular iff it is of the form $f^{-1}(M)$ for $M \subseteq S$, with S a finite semigroup.

Fix a nondeterministic automaton A of states Q

$$S = \mathcal{P}(Q \times Q) \qquad a.b = \{(p,r) \ : \ (p,q) \in a, \ (q,r) \in b\}$$

$$f(u) = \{(p,q) \ : \ p \longrightarrow_u q\} \qquad M = \mathit{Init} \times \mathit{Final}$$

SEMIGROUPS

Semigroup: $(S, .)$, with multiplication associative.
E.g., $(A^*, .)$ is a semigroup.
Morphism: $f : S \to S'$ with $f(x.y) = f(x).f(y)$

Th(Kleene): A language is regular iff it is of the form $f^{-1}(M)$ for $M \subseteq S$, with S a finite semigroup.

Fix a nondeterministic automaton A of states Q

$$S = \mathcal{P}(Q \times Q) \qquad a.b = \{(p,r) \ : \ (p,q) \in a, \ (q,r) \in b\}$$

$$f(u) = \{(p,q) \ : \ p \longrightarrow_u q\} \qquad M = \mathit{Init} \times \mathit{Final}$$

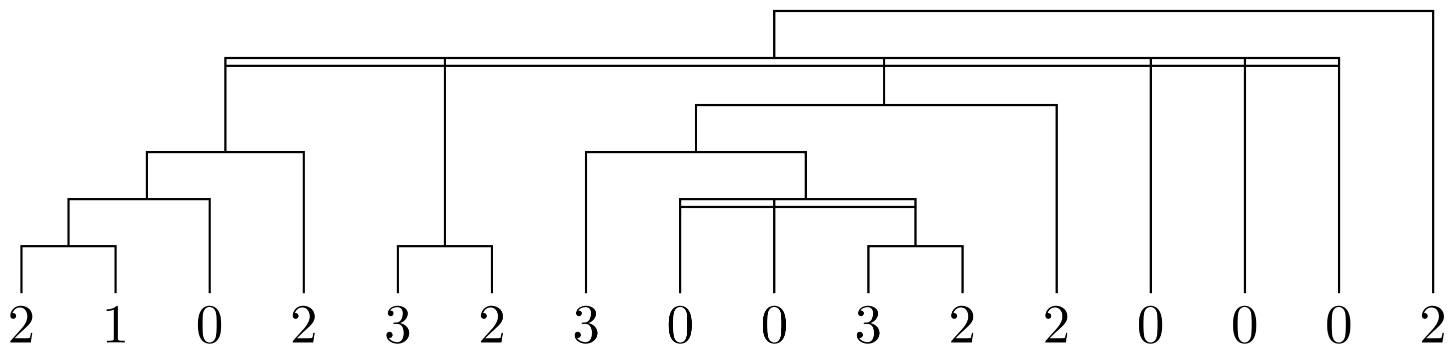
Below we identify S with the alphabet (f is the identity).
I.e., we work with words in S^*.
Call $f(u)$ the value of u.

Factorisation theorem by an example

Set $S = (\mathbb{Z}/5\mathbb{Z}, +)$ and the word $u =$210232300322002.

Factorisation theorem by an example

Set $S = (\mathbb{Z}/5\mathbb{Z}, +)$ and the word $u =$210232300322002.

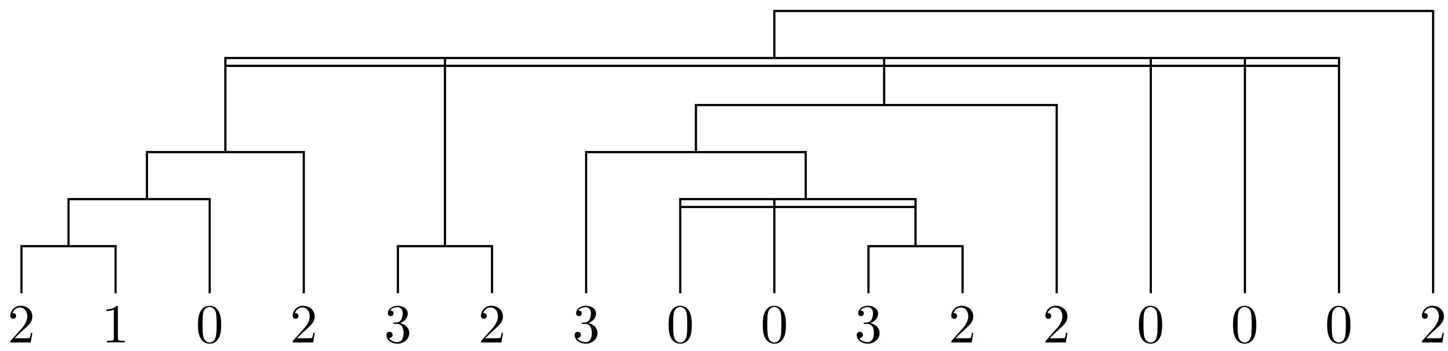


A factorising tree is a tree such that:

- leaves are labeled by letters,
- reading leaves from left to right yields the word.

FACTORISATION THEOREM BY AN EXAMPLE

Set $S = (\mathbb{Z}/5\mathbb{Z}, +)$ and the word $u =$210232300322002.



A factorising tree is a tree such that:

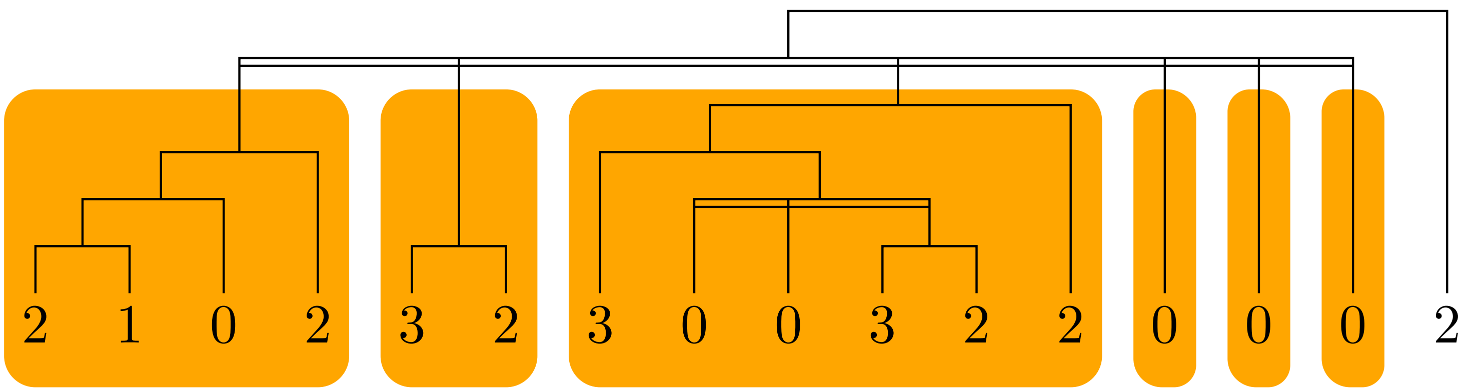
- leaves are labeled by letters,
- reading leaves from left to right yields the word.

A factorising tree is Ramseyan if every node:

- is a leaf, or;
- has two children, or;
- all its children have as value the same idempotent of S.

FACTORISATION THEOREM BY AN EXAMPLE

Set $S = (\mathbb{Z}/5\mathbb{Z}, +)$ and the word $u =$210232300322002.



A factorising tree is a tree such that:

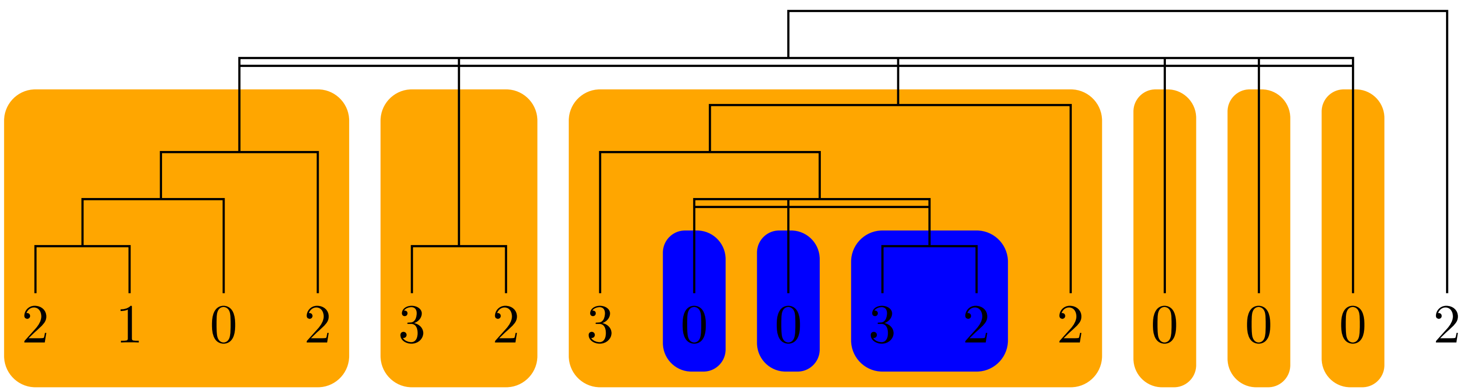
- leaves are labeled by letters,
- reading leaves from left to right yields the word.

A factorising tree is Ramseyan if every node:

- is a leaf, or;
- has two children, or;
- all its children have as value the same idempotent of S.

FACTORISATION THEOREM BY AN EXAMPLE

Set $S = (\mathbb{Z}/5\mathbb{Z}, +)$ and the word u =210232300322002.



A factorising tree is a tree such that:

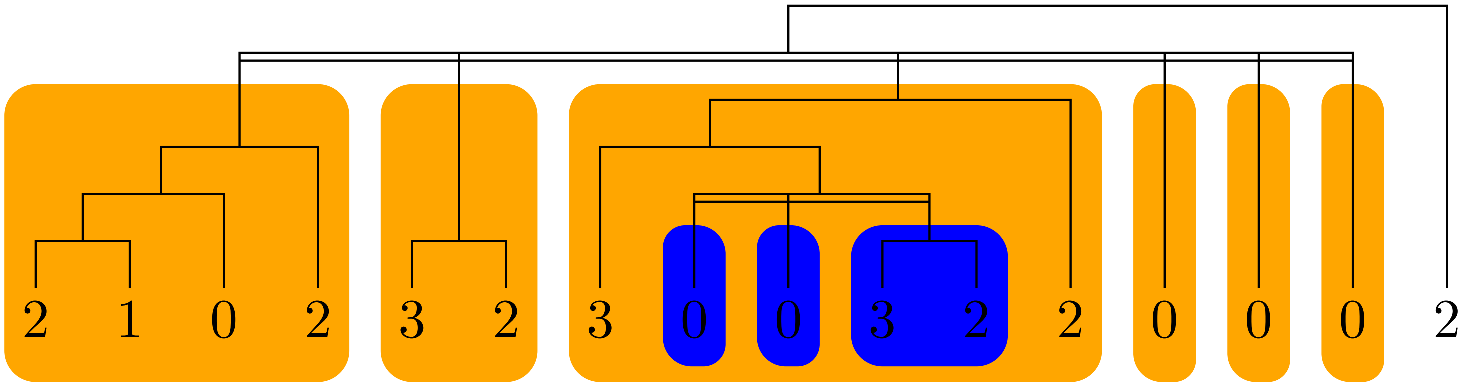
- leaves are labeled by letters,
- reading leaves from left to right yields the word.

A factorising tree is Ramseyan if every node:

- is a leaf, or;
- has two children, or;
- all its children have as value the same idempotent of S.

FACTORISATION THEOREM BY AN EXAMPLE

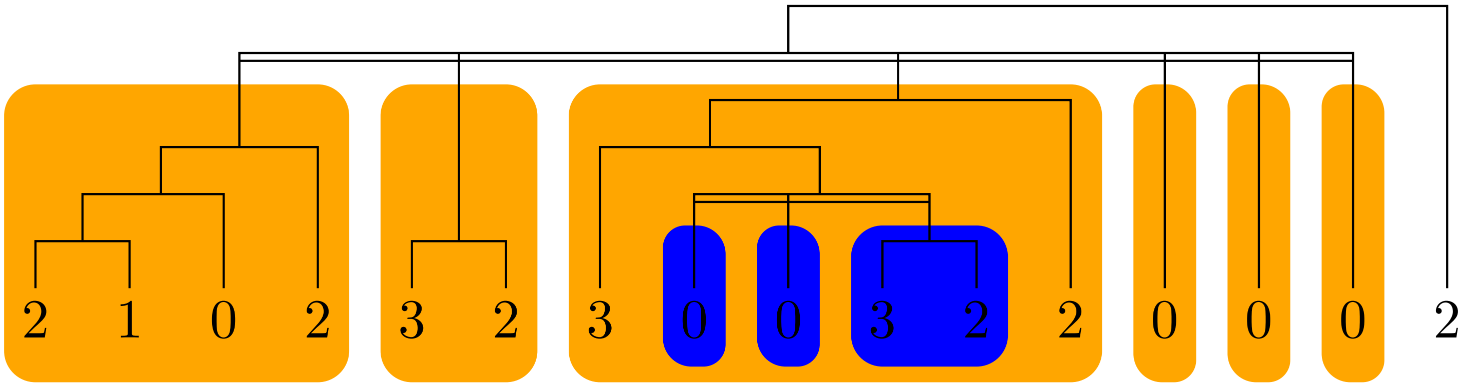
Set $S = (\mathbb{Z}/5\mathbb{Z}, +)$ and the word u =210232300322002.



Th(Simon 90): Every word over a finite semigroup S admits a Ramseyan factorisation tree of height at most $9|S|$.

FACTORISATION THEOREM BY AN EXAMPLE

Set $S = (\mathbb{Z}/5\mathbb{Z}, +)$ and the word $u =$210232300322002.

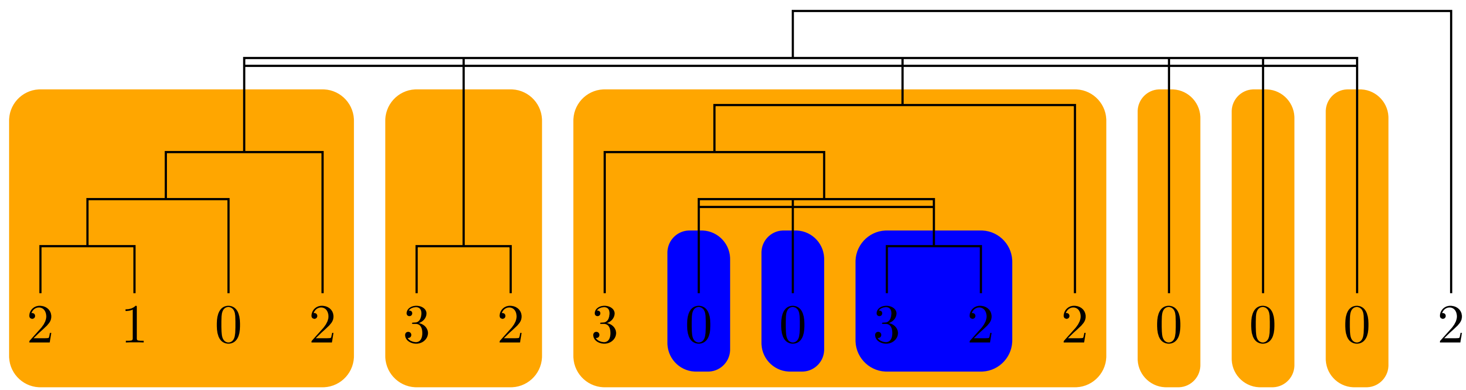


Th(Simon 90): Every word over a finite semigroup S admits a Ramseyan factorisation tree of height at most $9|S|$.

Th(Chalopin&Leung 04): Every word over a finite semigroup S admits a Ramseyan factorisation tree of height at most $7|S|$.

FACTORISATION THEOREM BY AN EXAMPLE

Set $S = (\mathbb{Z}/5\mathbb{Z}, +)$ and the word $u =$210232300322002.



Th(Simon 90): Every word over a finite semigroup S admits a Ramseyan factorisation tree of height at most $9|S|$.

Th(Chalopin&Leung 04): Every word over a finite semigroup S admits a Ramseyan factorisation tree of height at most $7|S|$.

Th: Every word over a finite semigroup S admits a Ramseyan factorisation tree of height at most $3|S|$.

Statement by regular expression

Th: For every finite semigroup S, there exists a regular expression such that:

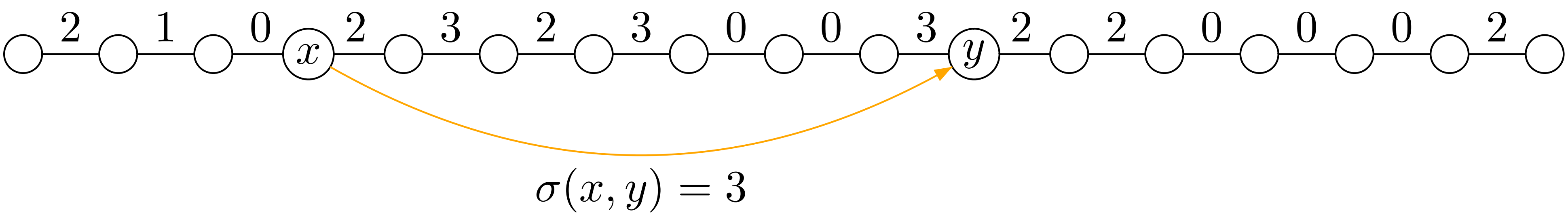
- it evaluates to S^*
- Kleene star L^* is allowed only if all the values of words in L are equal to the same idempotent

E.g., $S = (\mathbb{Z}/2\mathbb{Z}, +)$

$$0^*(10^*10^*)^*.(\varepsilon + 10^*)$$

Statement by splits

$S = (\mathbb{Z}/5\mathbb{Z})$

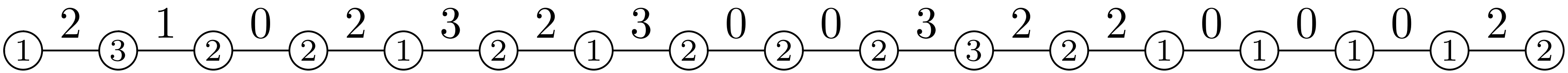


Def: Let α be a linear ordering, an additive labeling $\sigma : \alpha^2 \to S$ is such that:

- $\sigma(x,y)$ is defined iff $x < y$, and,
- $\forall x < y < z \in \alpha, \qquad \sigma(x,z) = \sigma(x,y).\sigma(y,z)$

Statement by splits

$S = (\mathbb{Z}/5\mathbb{Z})$



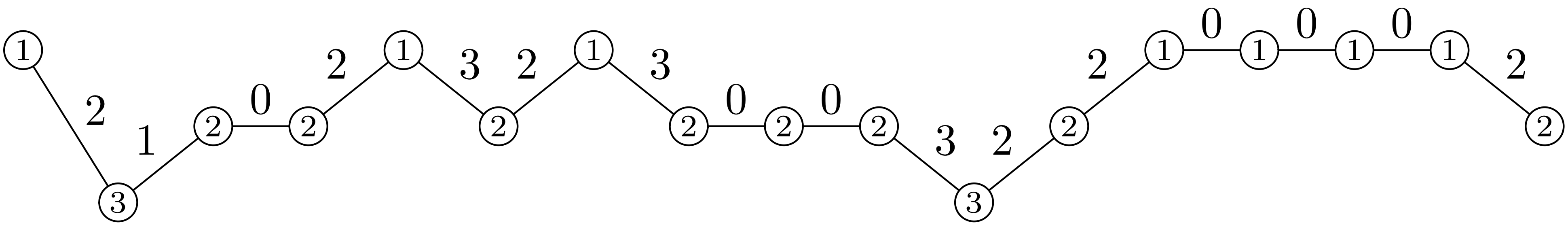
Def: Let α be a linear ordering, an additive labeling $\sigma : \alpha^2 \to S$ is such that:

- $\sigma(x, y)$ is defined iff $x < y$, and,
- $\forall x < y < z \in \alpha, \qquad \sigma(x, z) = \sigma(x, y).\sigma(y, z)$

Def: A split of height N is a mapping $s : \alpha \to [1, N]$.

STATEMENT BY SPLITS

$S = (\mathbb{Z}/5\mathbb{Z})$



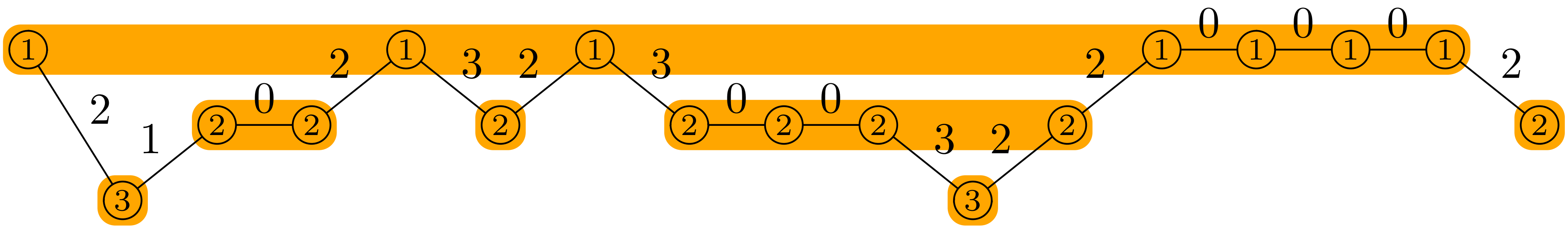
Def: Let α be a linear ordering, an additive labeling $\sigma : \alpha^2 \to S$ is such that:

- $\sigma(x, y)$ is defined iff $x < y$, and,
- $\forall x < y < z \in \alpha, \qquad \sigma(x, z) = \sigma(x, y).\sigma(y, z)$

Def: A split of height N is a mapping $s : \alpha \to [1, N]$.

Statement by splits

$S = (\mathbb{Z}/5\mathbb{Z})$



Def: Let α be a linear ordering, an additive labeling $\sigma : \alpha^2 \to S$ is such that:

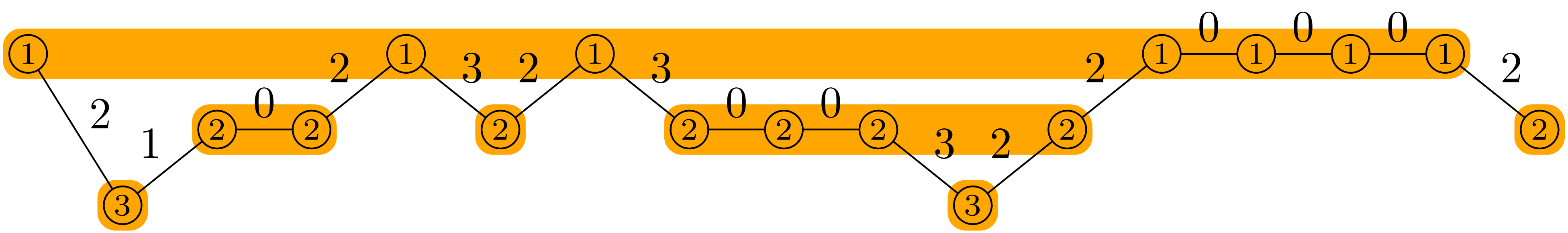
- $\sigma(x, y)$ is defined iff $x < y$, and,
- $\forall x < y < z \in \alpha, \qquad \sigma(x, z) = \sigma(x, y).\sigma(y, z)$

Def: A split of height N is a mapping $s : \alpha \to [1, N]$.

Def: $x \sim_s y$ si $s(x) = s(y)$ and for all $z \in [\min(x, y), \max(x, y)]$, $s(z) \geq s(x)$.

Statement by splits

$S = (\mathbb{Z}/5\mathbb{Z})$



Def: Let α be a linear ordering, an additive labeling $\sigma : \alpha^2 \to S$ is such that:

- $\sigma(x, y)$ is defined iff $x < y$, and,
- $\forall x < y < z \in \alpha, \qquad \sigma(x, z) = \sigma(x, y).\sigma(y, z)$

Def: A split of height N is a mapping $s : \alpha \to [1, N]$.

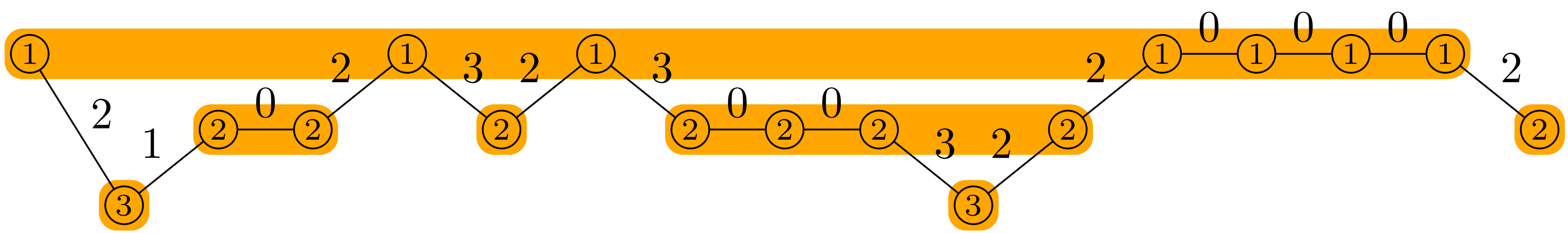
Def: $x \sim_s y$ si $s(x) = s(y)$ and for all $z \in [\min(x, y), \max(x, y)]$, $s(z) \geq s(x)$.

Def: A split is Ramseyan for σ if:

$$\forall x < y, x' < y'. \quad (x \sim_s y \sim_s x' \sim_s y') \to \sigma(x, y) = \sigma(x, y)^2 = \sigma(x', y')$$

Statement by splits

$S = (\mathbb{Z}/5\mathbb{Z})$



Def: Let α be a linear ordering, an additive labeling $\sigma : \alpha^2 \to S$ is such that:

- $\sigma(x, y)$ is defined iff $x < y$, and,
- $\forall x < y < z \in \alpha, \qquad \sigma(x, z) = \sigma(x, y).\sigma(y, z)$

Def: A split of height N is a mapping $s : \alpha \to [1, N]$.

Def: $x \sim_s y$ si $s(x) = s(y)$ and for all $z \in [\min(x, y), \max(x, y)]$, $s(z) \geq s(x)$.

Def: A split is Ramseyan for σ if:

$$\forall x < y, x' < y'. \quad (x \sim_s y \sim_s x' \sim_s y') \to \sigma(x, y) = \sigma(x, y)^2 = \sigma(x', y')$$

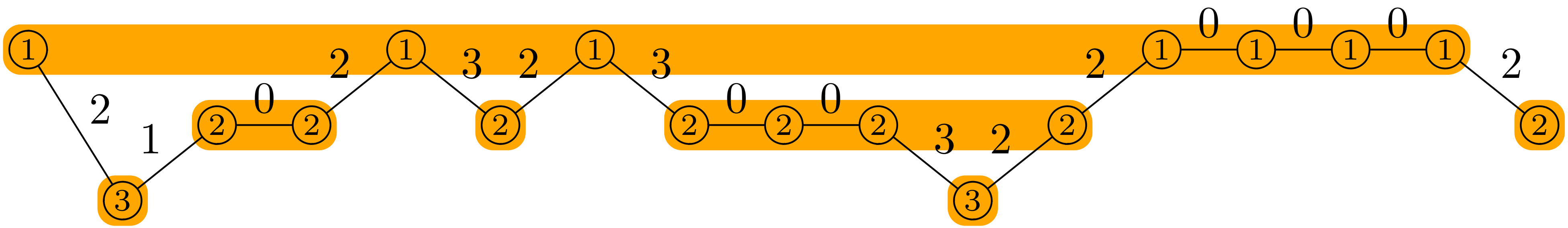
Th: Every additive labeling of a finite linear ordering by a finite semigroup S admits a Ramseyan split of height at most $|S|$.

An elementary application

Problem: Given a regular language L, and a word u, preprocess in linear time the word such that membership of a factor of u in L is constant time.

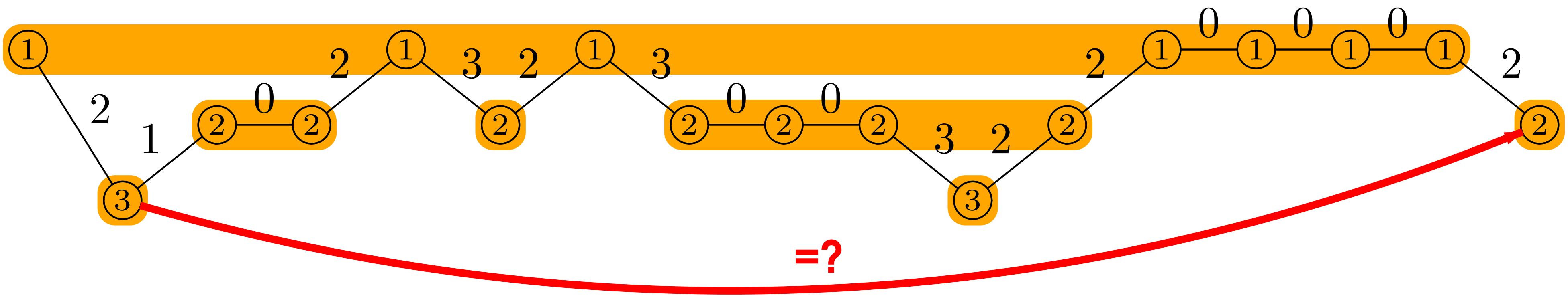
An elementary application

Problem: Given a regular language L, and a word u, preprocess in linear time the word such that membership of a factor of u in L is constant time.



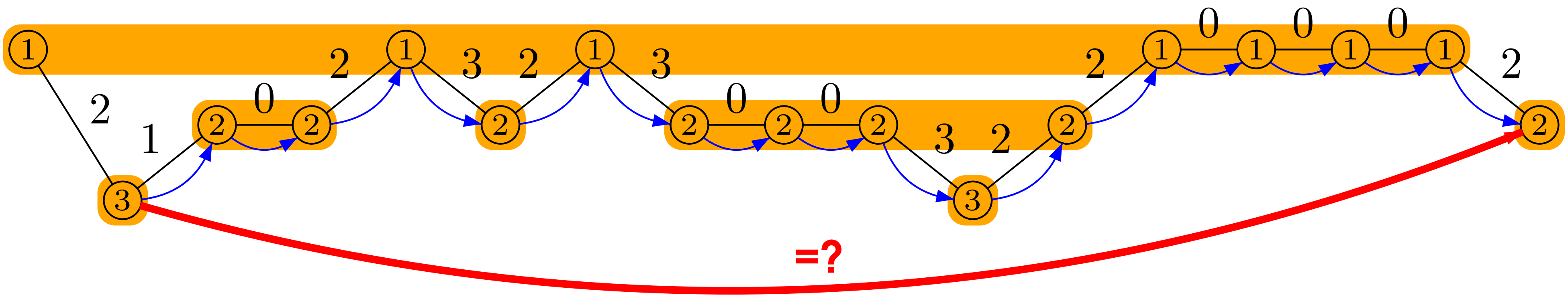
AN ELEMENTARY APPLICATION

Problem: Given a regular language L, and a word u, preprocess in linear time the word such that membership of a factor of u in L is constant time.



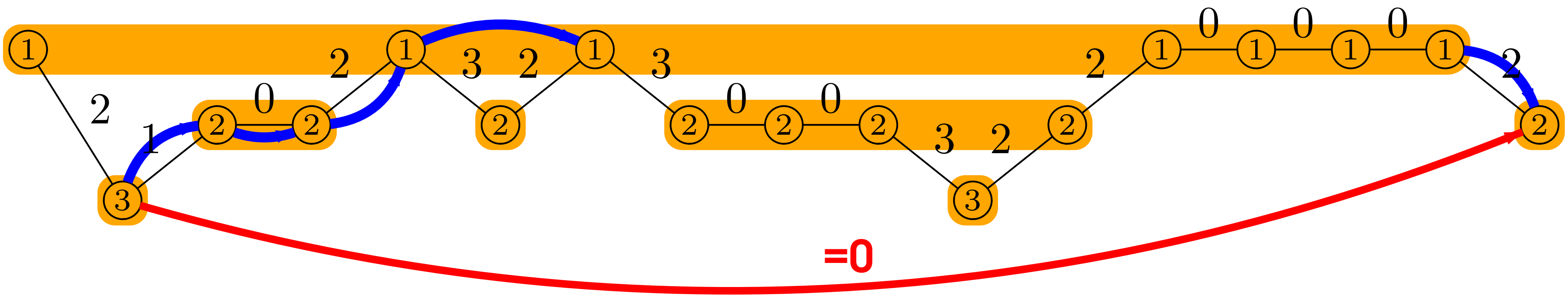
AN ELEMENTARY APPLICATION

Problem: Given a regular language L, and a word u, preprocess in linear time the word such that membership of a factor of u in L is constant time.



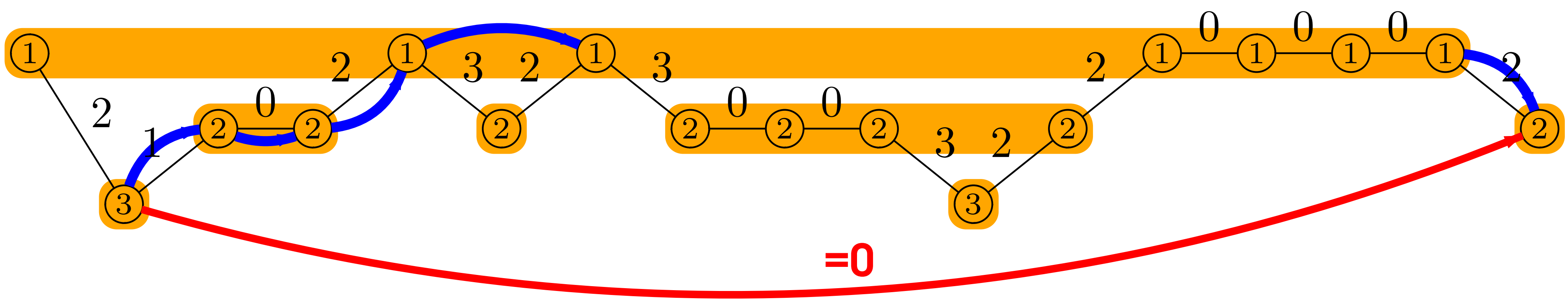
An elementary application

Problem: Given a regular language L, and a word u, preprocess in linear time the word such that membership of a factor of u in L is constant time.



An elementary application

Problem: Given a regular language L, and a word u, preprocess in linear time the word such that membership of a factor of u in L is constant time.



Conclusion: Computing $\sigma(x, y)$ for a given x, y is constant.

SOME ARGUMENTS OF THE PROOF

Situation 1: $(S, .)$ is a group
every element has an inverse

Situation 2: $ab = aa = a$, $ba = bb = b$
every product collapse to its first factor

Situation 3: $aa = ba = a$, $ab = bb = b$
every product collapse to its last factor

Situation 4: $aa = ab = ba = a$, $bb = b$
An element can swallow another

Some arguments of the proof

Situation 1: $(S, .)$ is a group
every element has an inverse
Single regular $\mathcal{H}$-class

Situation 2: $ab = aa = a$, $ba = bb = b$
every product collapse to its first factor
Single regular $\mathcal{L}$-class

Situation 3: $aa = ba = a$, $ab = bb = b$
every product collapse to its last factor
Single regular $\mathcal{R}$-class

Situation 4: $aa = ab = ba = a$, $bb = b$
An element can swallow another
Multiple $\mathcal{J}$-classes

Some arguments of the proof

Situation 1: $(S, .)$ is a group
every element has an inverse
Single regular $\mathcal{H}$-class

Situation 2: $ab = aa = a$, $ba = bb = b$
every product collapse to its first factor
Single regular $\mathcal{L}$-class

Situation 3: $aa = ba = a$, $ab = bb = b$
every product collapse to its last factor
Single regular $\mathcal{R}$-class

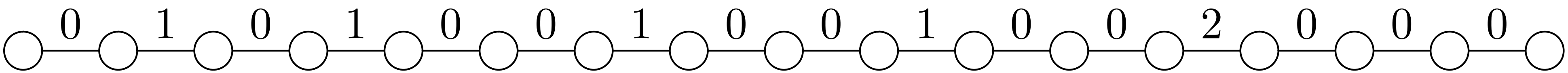
Situation 4: $aa = ab = ba = a$, $bb = b$
An element can swallow another
Multiple $\mathcal{J}$-classes

The proof uses a different argument in each situation.

Those arguments are combined in a proof for every finite semigroup (Green's relations).

Some arguments of the proof

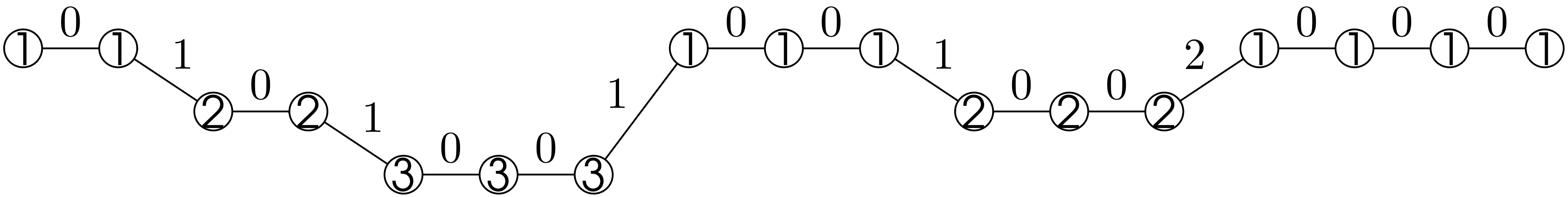
Group case. E.g., $(\mathbb{Z}/3\mathbb{Z}, +)$



Number S by n. Set $s(x) = n(\sigma(0, x))$, $s(0) = n(e)$.

Some arguments of the proof

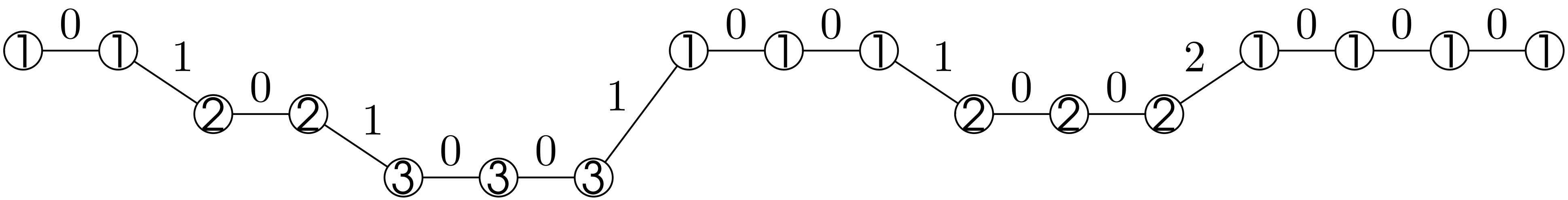
Group case. E.g., $(\mathbb{Z}/3\mathbb{Z}, +)$



Number S by n. Set $s(x) = n(\sigma(0, x))$, $s(0) = n(e)$.

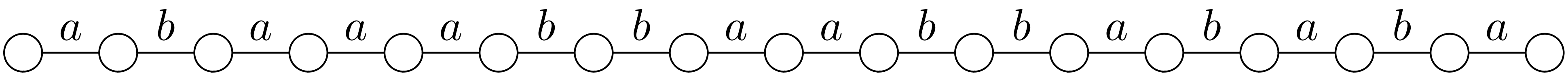
SOME ARGUMENTS OF THE PROOF

Group case. E.g., $(\mathbb{Z}/3\mathbb{Z}, +)$



Number S by n. Set $s(x) = n(\sigma(0, x))$, $s(0) = n(e)$.

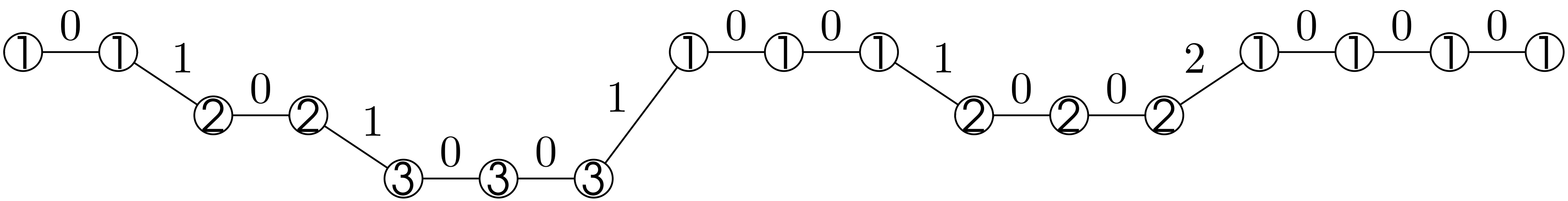
$\mathcal{L}$-equivalent elements. $ab = bb = b$, $aa = ba = a$.



Number S by $n : S \to [1, N]$. Set $s(x) = n(\sigma(x - 1, x))$, $s(0) = 1$.

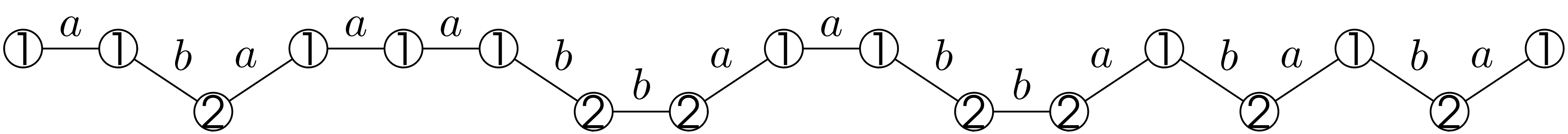
Some arguments of the proof

Group case. E.g., $(\mathbb{Z}/3\mathbb{Z},+)$



Number S by n. Set $s(x) = n(\sigma(0,x))$, $s(0) = n(e)$.

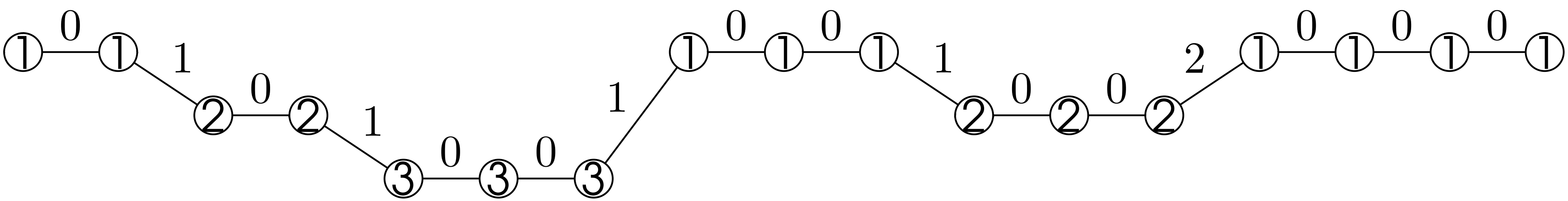
$\mathcal{L}$-equivalent elements. $ab = bb = b$, $aa = ba = a$.



Number S by $n : S \to [1,N]$. Set $s(x) = n(\sigma(x-1,x))$, $s(0) = 1$.

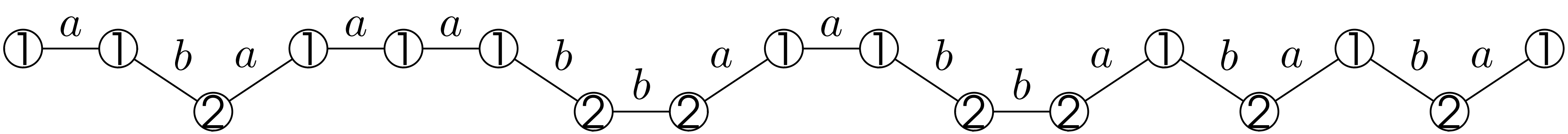
Some arguments of the proof

Group case. E.g., $(\mathbb{Z}/3\mathbb{Z}, +)$



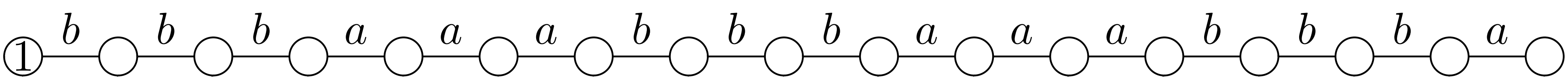
Number S by n. Set $s(x) = n(\sigma(0,x))$, $s(0) = n(e)$.

$\mathcal{L}$-equivalent elements. $ab = bb = b$, $aa = ba = a$.



Number S by $n : S \to [1, N]$. Set $s(x) = n(\sigma(x-1,x))$, $s(0) = 1$.

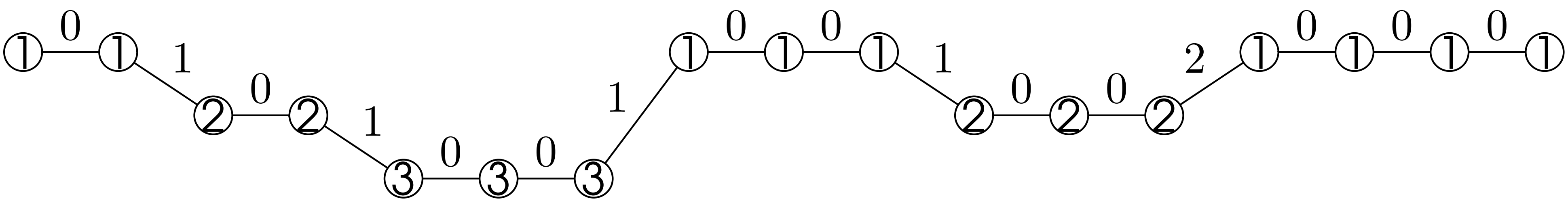
Case of different $\mathcal{J}$-classes. $c = *c = c* = ab \quad \leq_{\mathcal{J}} \quad b = bb = ba \quad \leq_{\mathcal{J}} \quad a = aa$



Fix a minimal remaining $\mathcal{J}$-class. Cut at the first occurence of this class. Proceed.

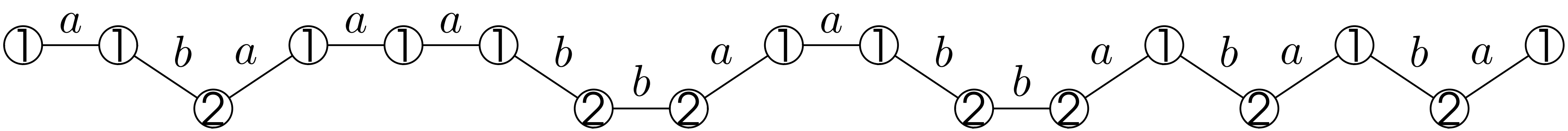
Some arguments of the proof

Group case. E.g., $(\mathbb{Z}/3\mathbb{Z}, +)$



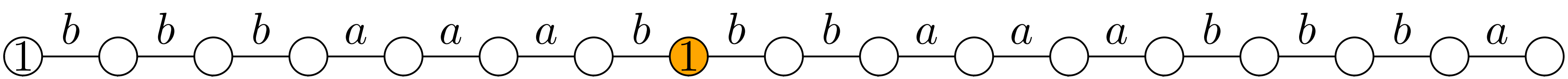
Number S by n. Set $s(x) = n(\sigma(0, x))$, $s(0) = n(e)$.

$\mathcal{L}$-equivalent elements. $ab = bb = b$, $aa = ba = a$.



Number S by $n : S \to [1, N]$. Set $s(x) = n(\sigma(x-1, x))$, $s(0) = 1$.

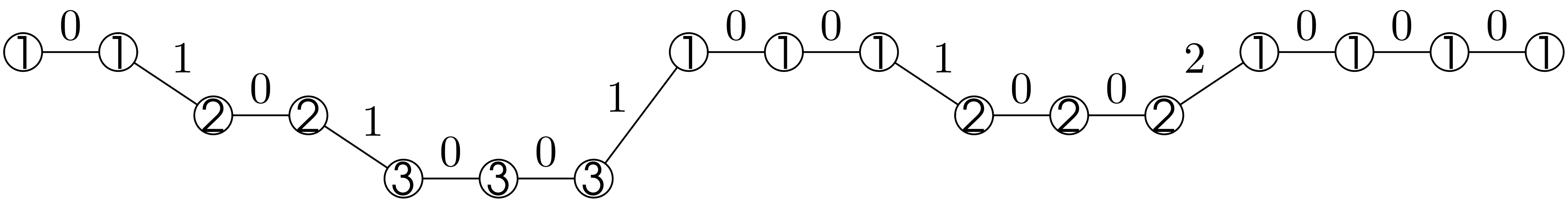
Case of different $\mathcal{J}$-classes. $c = *c = c* = ab \quad \leq_{\mathcal{J}} \quad b = bb = ba \quad \leq_{\mathcal{J}} \quad a = aa$



Fix a minimal remaining $\mathcal{J}$-class. Cut at the first occurence of this class. Proceed.

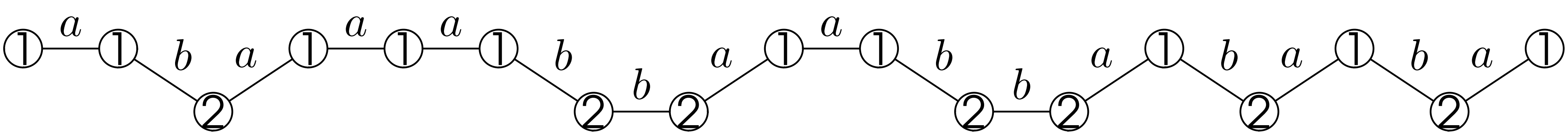
Some arguments of the proof

Group case. E.g., $(\mathbb{Z}/3\mathbb{Z}, +)$



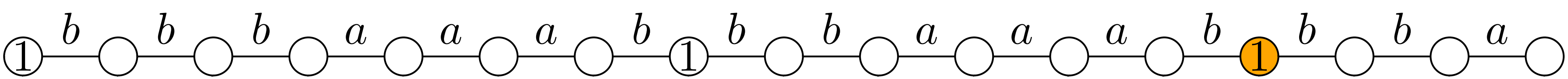
Number S by n. Set $s(x) = n(\sigma(0, x))$, $s(0) = n(e)$.

$\mathcal{L}$-equivalent elements. $ab = bb = b$, $aa = ba = a$.



Number S by $n : S \to [1, N]$. Set $s(x) = n(\sigma(x-1, x))$, $s(0) = 1$.

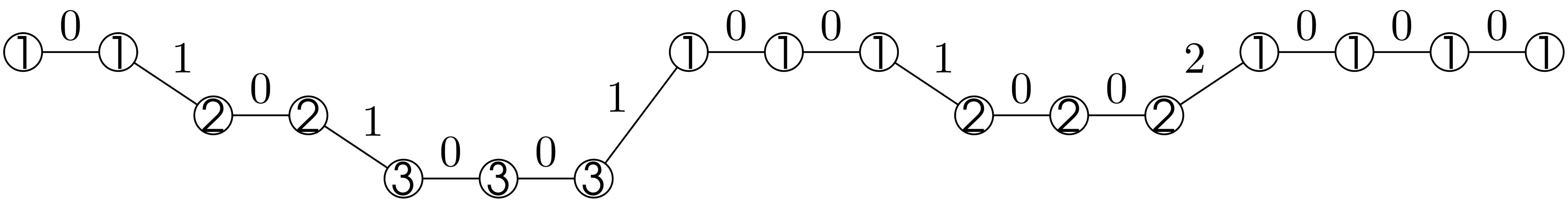
Case of different $\mathcal{J}$-classes. $c = *c = c* = ab \quad \leq_{\mathcal{J}} \quad b = bb = ba \quad \leq_{\mathcal{J}} \quad a = aa$



Fix a minimal remaining $\mathcal{J}$-class. Cut at the first occurence of this class. Proceed.

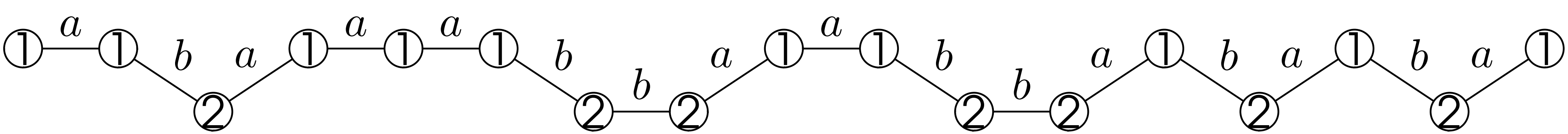
Some arguments of the proof

Group case. E.g., $(\mathbb{Z}/3\mathbb{Z}, +)$



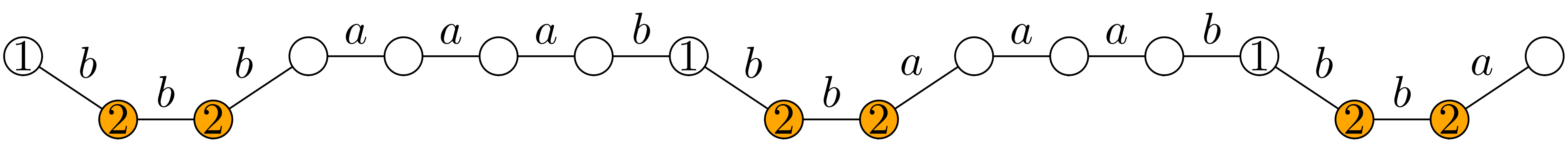
Number S by n. Set $s(x) = n(\sigma(0,x))$, $s(0) = n(e)$.

$\mathcal{L}$-equivalent elements. $ab = bb = b$, $aa = ba = a$.



Number S by $n : S \to [1, N]$. Set $s(x) = n(\sigma(x-1,x))$, $s(0) = 1$.

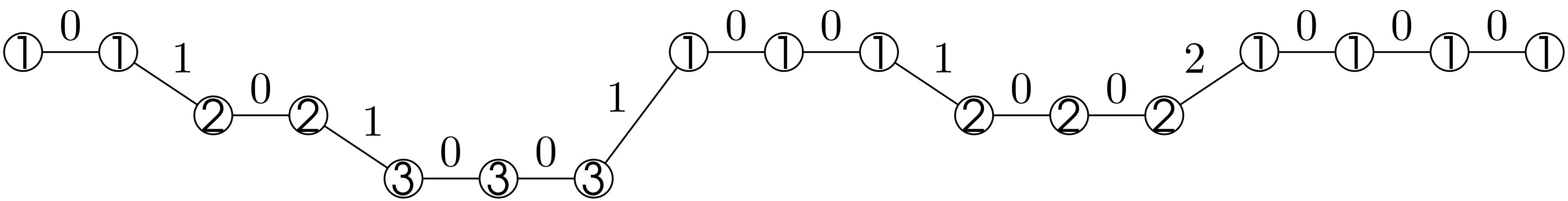
Case of different $\mathcal{J}$-classes. $c = *c = c* = ab \quad \leq_{\mathcal{J}} \quad b = bb = ba \quad \leq_{\mathcal{J}} \quad a = aa$



Fix a minimal remaining $\mathcal{J}$-class. Cut at the first occurence of this class. Proceed.

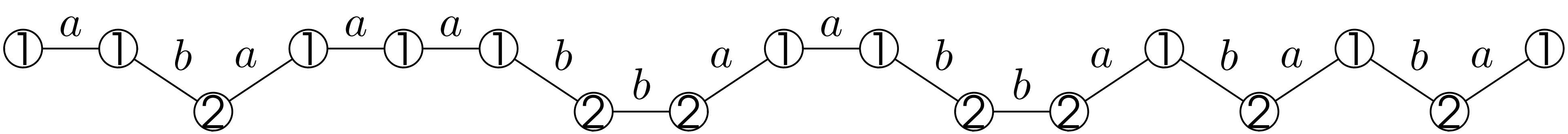
Some arguments of the proof

Group case. E.g., $(\mathbb{Z}/3\mathbb{Z}, +)$



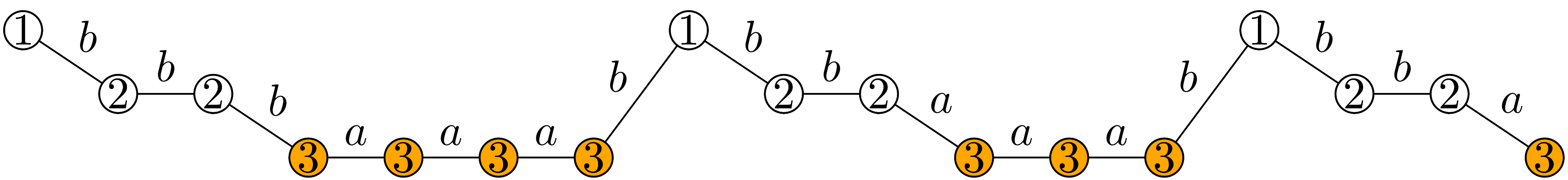
Number S by n. Set $s(x) = n(\sigma(0,x))$, $s(0) = n(e)$.

$\mathcal{L}$-equivalent elements. $ab = bb = b$, $aa = ba = a$.



Number S by $n : S \to [1, N]$. Set $s(x) = n(\sigma(x-1,x))$, $s(0) = 1$.

Case of different $\mathcal{J}$-classes. $c = {*c} = c* = ab \quad \leq_{\mathcal{J}} \quad b = bb = ba \quad \leq_{\mathcal{J}} \quad a = aa$



Fix a minimal remaining $\mathcal{J}$-class. Cut at the first occurence of this class. Proceed.

EXTENSION TO INFINITE LINEAR ORDERINGS

Th: Every additive labeling of a finite linear ordering by a finite semigroup S admits a Ramseyan split of height at most $|S|$.

Th: Every additive labeling of a complete linear ordering by a finite semigroup S admits a Ramseyan split of height at most $3|S|$.

SCATTERED LINEAR ORDERINGS AND AUTOMATA

Buchi introduces automata accepting words indexed by ω.

Scattered linear orderings and automata

Buchi introduces automata accepting words indexed by ω.

Th(Buchi 61): Languages accepted by Buchi automata are closed under union, intersection, projection and complementation.

Key of the proof: Uses Ramsey's theorem to uncover regularity in the behaviour of the automaton aver a given word.

Scattered linear orderings and automata

Buchi introduces automata accepting words indexed by ω.

Th(Buchi 61): Languages accepted by Buchi automata are closed under union, intersection, projection and complementation.

Key of the proof: Uses Ramsey's theorem to uncover regularity in the behaviour of the automaton aver a given word.

Def: A linear ordering is scattered if it does not contain $(\mathbb{Q}, <)$ as a subordering.
E.g., Every ordinal is scattered
E.g., $(\mathbb{Z}, <)$ is scattered, but $(\mathbb{R}, <)$ is not

Bruyère & Carton introduce a family of automata for words indexed by countable scattered linear orderings.

SCATTERED LINEAR ORDERINGS AND AUTOMATA

Buchi introduces automata accepting words indexed by ω.

Th(Buchi 61): Languages accepted by Buchi automata are closed under union, intersection, projection and complementation.

Key of the proof: Uses Ramsey's theorem to uncover regularity in the behaviour of the automaton aver a given word.

Def: A linear ordering is scattered if it does not contain $(\mathbb{Q}, <)$ as a subordering.
E.g., Every ordinal is scattered
E.g., $(\mathbb{Z}, <)$ is scattered, but $(\mathbb{R}, <)$ is not

Bruyère & Carton introduce a family of automata for words indexed by countable scattered linear orderings.

Th(Carton&Rispal 05): Languages accepted by Bruyère and Carton automata are closed under union, intersection, projection and complementation.

Scattered linear orderings and automata

Buchi introduces automata accepting words indexed by ω.

Th(Buchi 61): Languages accepted by Buchi automata are closed under union, intersection, projection and complementation.

Key of the proof: Uses Ramsey's theorem to uncover regularity in the behaviour of the automaton aver a given word.

Def: A linear ordering is scattered if it does not contain $(\mathbb{Q}, <)$ as a subordering.
E.g., Every ordinal is scattered
E.g., $(\mathbb{Z}, <)$ is scattered, but $(\mathbb{R}, <)$ is not

Bruyère & Carton introduce a family of automata for words indexed by countable scattered linear orderings.

Th(Carton&Rispal 05): Languages accepted by Bruyère and Carton automata are closed under union, intersection, projection and complementation.

Contribution: A simpler proof using splits of complete linear orderings.

RELATED WORK

Other uses of Simon's factorisation theorem

- caracterisation of subfamilies of regular languages (Pin&Weil)
- limitedness of distance automata (Simon)
- extension of regularity over ω-words (Bojańczyk&C.)

Variants

- Deterministic variant, application to logic over trees (Icalp 07)