

Profinite topologies on words

Jean-Éric Pin

LIAFA, CNRS and University Paris 7

10 January 2007, Buenos-Aires



Outline

- (1) Preliminaries
- (2) Profinite topologies
- (3) Equational definitions of varieties
- (4) The p -adic case
- (5) The group case
- (6) Applications to finite automata and semigroups
- (7) Conclusion and open problems

Semigroups, monoids and groups

A **semigroup** is a set equipped with an associative binary operation. A **monoid** is a semigroup with a unit element. A **group** is a monoid in which each element has an inverse.

Let p be a prime number. A **p -group** is a finite group of size p^n for some $n > 0$.

Given an alphabet A , one denotes by A^* [A^+ , $FG(A)$] the **free monoid** [**semigroup**, **group**] on A . The **empty word** is denoted by 1 .

Varieties

A **Birkhoff variety of semigroups** is a class of semigroups closed under taking subsemigroups, quotients (= homomorphic images) and direct products.

A **variety of finite semigroups** is a class of **finite** semigroups closed under taking subsemigroups, quotients and **finite** direct products.

- Similar definitions can be given for **monoids** and for **groups**.
- Finite **groups** form a variety of finite monoids.

Part I

Profinite topologies



Separating words

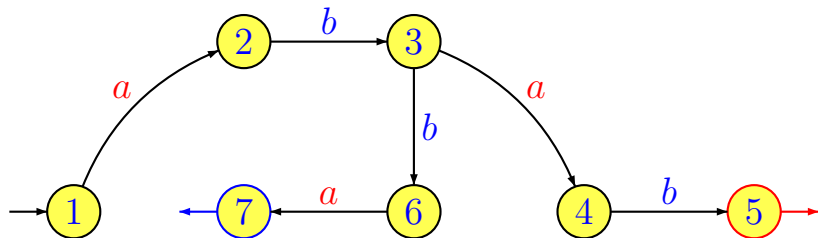
A monoid M separates two words u and v of A^* if there exists a monoid morphism $\varphi : A^* \rightarrow M$ such that $\varphi(u) \neq \varphi(v)$.

For instance, the morphism which maps each word onto its length modulo 2 is a morphism from $\{a, b\}^*$ onto $\mathbb{Z}/2\mathbb{Z}$ which separates $abaaba$ and $abaabab$.

Let $M = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\}$ and let $\varphi : \{a, b\}^* \rightarrow M$ defined by $\varphi(a) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ and $\varphi(b) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$. Then, for each word u , φ separates ua and ub since $\varphi(ua) = \varphi(a)$ and $\varphi(ub) = \varphi(b)$.

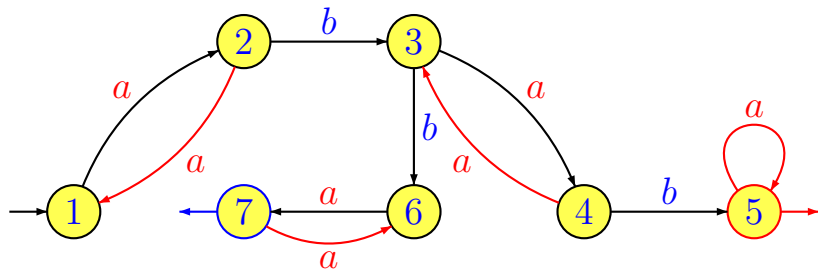
Separating *abab* and *abba* with a finite group

(1) Build an automaton with the two words



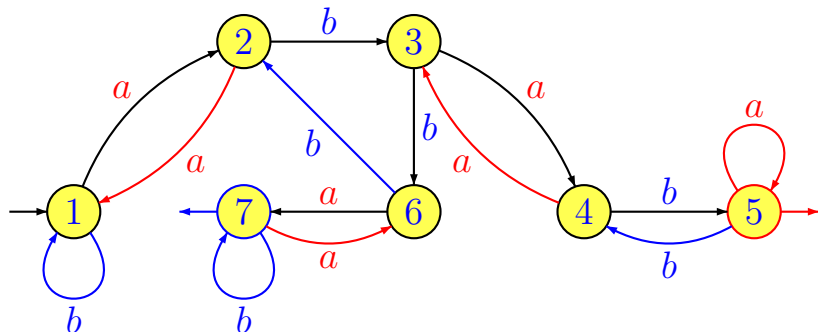
Separating *abab* and *abba* with a finite group

(1) Build an automaton with the two words



Separating *abab* and *abba* with a finite group

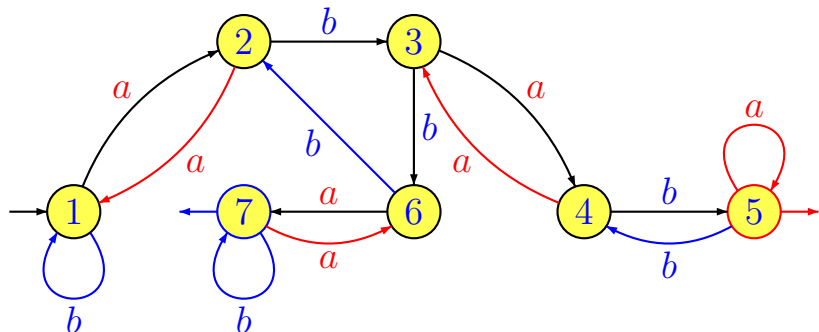
(1) Build an automaton with the two words



(2) Complete into **permutations**.

Separating $abab$ and $abba$ with a finite group

(1) Build an automaton with the two words



(2) Complete into **permutations**. The resulting permutation group separates $abab$ and $abba$ since $1 \cdot abab = 5$ and $1 \cdot abba = 7$.

Some metrics on words

Let u and v be two words. Put

$$r(u, v) = \min \{ |M| \mid M \text{ is a finite monoid} \\ \text{that separates } u \text{ and } v \}$$

$$r_{\mathbf{G}}(u, v) = \min \{ |G| \mid G \text{ is a finite group} \\ \text{that separates } u \text{ and } v \}$$

$$d(u, v) = 2^{-r(u, v)} \quad d_{\mathbf{G}}(u, v) = 2^{-r_{\mathbf{G}}(u, v)}$$

- Intuitively, two words are close for d [$d_{\mathbf{G}}$] if one needs a **large** monoid [group] to separate them.
- One could give a similar definition for any **variety of finite monoids**.



Main properties of d and $d_{\mathbf{G}}$

For all $u, v, w \in A^*$,

- $d(u, v) = 0$ iff $u = v$,
- $d(u, v) \leq \max\{d(u, w), d(w, v)\}$ and hence d [$d_{\mathbf{G}}$] is an ultrametric,
- $d(uw, vw) \leq d(u, v)$, $d(wu, wv) \leq d(u, v)$,
- the product of words is uniformly continuous,
- any monoid morphism from A^* to B^* is uniformly continuous for d [$d_{\mathbf{G}}$],
- any monoid morphism from (A^*, d) [$(A^*, d_{\mathbf{G}})$] onto a finite discrete monoid [group] is uniformly continuous.

Main properties of d and $d_{\mathbf{G}}$ (continued)

- A sequence of words $(u_n)_{n \geq 0}$ is a **Cauchy sequence** for d [for $d_{\mathbf{G}}$] iff, for every monoid morphism φ from A^* to a finite monoid [group], the sequence $\varphi(u_n)_{n \geq 0}$ is ultimately constant.
- A sequence of words $(u_n)_{n \geq 0}$ is **converging** to a word u (for d [for $d_{\mathbf{G}}$]) iff, for every monoid morphism φ from A^* to a finite monoid [group], the sequence $\varphi(u_n)_{n \geq 0}$ is ultimately equal to $\varphi(u)$.

A converging sequence for d_G

Theorem (M. Hall 1950)

For each word x , the sequence $(x^{n!})_{n \geq 0}$ converges to the empty word in the pro-group topology.

Proof. Let $\varphi : A^* \rightarrow G$ be a monoid morphism onto a finite group G and let $g = \varphi(x)$. We claim that the sequence $\varphi(x^{n!}) = g^{n!}$ is ultimately equal to $\varphi(1) = 1$.

By Lagrange's theorem, $g^{|G|} = 1$. Now, for $n \geq |G|$, $|G|$ divides $n!$ and hence, $g^{n!} = 1$. \square

Free profinite monoids and groups

The metric space (A^*, d) $[(A^*, d_G)]$ is not complete. Its completion is the free **profinite monoid** [**profinite group**] on A and is denoted by $\widehat{A^*}$ [$\widehat{FG(A)}$]. If A is **finite**, the completion is **compact**.

The product is uniformly continuous on A^* and hence can be extended by continuity to the completion of A^* .

For each $x \in A^*$, the sequence $x^{n!-1}$ is a Cauchy sequence, and hence converges in $\widehat{FG(A)}$. Since

$$x \lim_{n \rightarrow \infty} x^{n!-1} = \lim_{n \rightarrow \infty} x^{n!-1} x = \lim_{n \rightarrow \infty} x^{n!} = 1$$

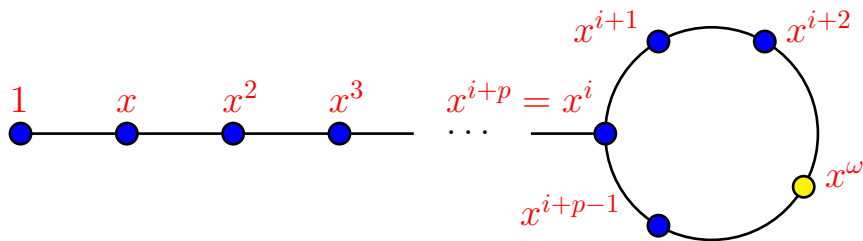
x has an inverse.



Properties of the profinite metric

Theorem (Reutenauer 1979)

For each $x \in A^*$, the sequence $x^{n!}$ converges in $\widehat{A^*}$ to a limit, denoted by x^ω .



Part II

Varieties

Quotation (M. Stone)

*A cardinal principle of modern mathematical research may be stated as a maxim: **One must always topologize.***



Identities

Let A be an alphabet and let $u, v \in A^+$. A semigroup S satisfies the identity $u = v$ iff, for each semigroup morphism $\varphi : A^+ \rightarrow S$, $\varphi(u) = \varphi(v)$.

- A semigroup satisfies the identity $xyx = x$ if, for each $x, y \in S$, $xyx = x$.
- A semigroup is **commutative** iff it satisfies the identity $xy = yx$.
- A semigroup is **idempotent** iff it satisfies the identity $x = x^2$.

Proposition (Easy!)

*Given a set E of identities, the class of all semigroups satisfying the identities of E is a **Birkhoff variety** of semigroups.*

Birkhoff's theorem

Proposition (Easy!)

*Given a set E of identities, the class of all semigroups satisfying the identities of E is a **Birkhoff variety** of semigroups.*

Theorem (Birkhoff 1935)

*A class of semigroups is a **Birkhoff variety** iff it is defined by a **set of identities**.*



Birkhoff's theorem

Proposition (Easy!)

*Given a set E of identities, the class of all semigroups satisfying the identities of E is a **Birkhoff variety** of semigroups.*

Theorem (Birkhoff 1935)

*A class of semigroups is a **Birkhoff variety** iff it is defined by a **set of identities**.*

What happens for finite semigroups ?



Reiterman's theorem

Define a **profinite identity** as a formal equality of the form $u = v$, where u and v are elements of a free profinite monoid.

Theorem (Reiterman 1982)

*A class of **finite** semigroups is a **variety** iff it is defined by a **set of profinite identities**.*

The variety of finite groups is defined by the single identity $x^\omega = 1$ since, in a finite group, the unique idempotent is the identity.



Part III

The pro- p topology

Let p be a prime number. Put

$$r_p(u, v) = \min\{|G| \mid G \text{ is a finite } p\text{-group} \\ \text{that separates } u \text{ and } v\}$$

$$d_p(u, v) = 2^{-r_p(u, v)}$$

The metric d_p defines the pro- p topology on A^* .

Subwords

A word $u = a_1a_2 \cdots a_k$ is a **subword** of v if $v = v_0a_1v_1a_2 \cdots a_kv_k$ for some $v_0, v_1, \dots, v_k \in A^*$.
For instance, **Bees** is a subword of **BuenosAires**.

Let $\binom{v}{u}$ be the number of distinct ways to write u as a subword of v . For instance,

$$\binom{abab}{ab} = 3 \text{ (} abab, abab, abab \text{)} \text{ and } \binom{a^n}{a^m} = \binom{n}{m}.$$

- For every word $u \in A^*$, $\binom{u}{1} = 1$.
- For every word $u \in A^+$, $\binom{1}{u} = 0$.
- If $w = uv$, then $\binom{w}{x} = \sum_{x_1x_2=x} \binom{u}{x_1} \binom{v}{x_2}$.

Computing the binomial coefficients modulo p

The function $\tau : A^* \rightarrow \mathcal{M}_{n+1}(\mathbb{Z}/p\mathbb{Z})$ defined by

$$\tau(u) = \begin{pmatrix} 1 & \binom{u}{a_1} & \binom{u}{a_1 a_2} & \binom{u}{a_1 a_2 a_3} & \cdots & \binom{u}{a_1 a_2 \cdots a_n} \\ 0 & 1 & \binom{u}{a_2} & \binom{u}{a_2 a_3} & \cdots & \binom{u}{a_2 \cdots a_n} \\ 0 & 0 & 1 & \binom{u}{a_3} & \cdots & \binom{u}{a_3 \cdots a_n} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \binom{u}{a_n} \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

is a morphism of monoids and the **unitriangular** $n \times n$ **matrices** with entries in $\mathbb{Z}/p\mathbb{Z}$ form a p -group.

Another metric for the pro- p topology

Let

$$r'_p(u, v) = \min \left\{ |x| \mid \binom{u}{x} \not\equiv \binom{v}{x} \pmod{p} \right\}$$

$$d'_p(u, v) = 2^{-r'_p(u, v)}$$

Then d'_p is a metric which also defines the pro- p topology on A^* .

Thue-Morse word

Let $A = \{a, b\}$, and let $\tau : A^* \rightarrow A^*$ be the monoid morphism defined by $\tau(a) = ab$ and $\tau(b) = ba$. The sequence $\tau^n(a)$ is $a, ab, abba, abbabaab, \dots$. It defines an infinite word $t = abbabaabbaababba \dots$ called the **Thue-Morse word**.

For each prefix $t[n]$ of t , we are interested in the value **mod** p of the numbers

$$\binom{t[n]}{a}, \binom{t[n]}{b}, \binom{t[n]}{aa}, \binom{t[n]}{ab}, \binom{t[n]}{ba}, \binom{t[n]}{bb}, \binom{t[n]}{aaa}, \binom{t[n]}{aab}, \dots$$

Some binomial coefficients $\binom{t[n]}{v}$

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
a	1	1	1	2	2	3	4	4	4	5	6	6	7	7	7	8
b	0	1	2	2	3	3	3	4	5	5	5	6	6	7	8	8
aa	0	0	0	1	1	3	6	6	6	10	15	15	21	21	21	28
ab	0	1	2	2	4	4	4	8	12	12	12	18	18	25	32	32
ba	0	0	0	2	2	5	8	8	8	13	18	18	24	24	24	32
bb	0	0	1	1	3	3	3	6	10	10	10	15	15	21	28	28

Some converging sequences

For every prime p and $n > 0$, let

$$f(p, n) = \begin{cases} 2^n p^{1+\lfloor \log_p n \rfloor} & \text{if } p \neq 2 \\ 2^k \text{ where } F_{k-1} \leq n < F_k & \text{if } p = 2 \end{cases}$$

Theorem

Let $m = f(p, n)$. Then for every non-empty word v of length $\leq n$, $\binom{m}{v} \equiv 0 \pmod{p}$. Thus the sequence $(t[f(p, n)])_{n>0}$ converges to the empty word for d_p .

Part IV

Back to the pro-group topology

Let L be a **regular language** of A^* (= set of words recognized by a finite automaton). Is it **decidable** whether L is **open**, **closed** or **clopen** in the pro-group topology?

Group languages

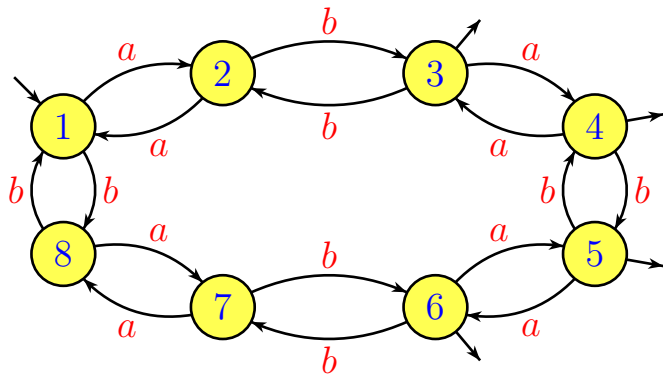
- A **group automaton** is a finite deterministic automaton in which each letter defines a **permutation** on the set of states.
- A **group language** is a language accepted by a **group automaton**.

Proposition

*A regular language is **clopen** in the pro-group topology iff it is a **group language**.*

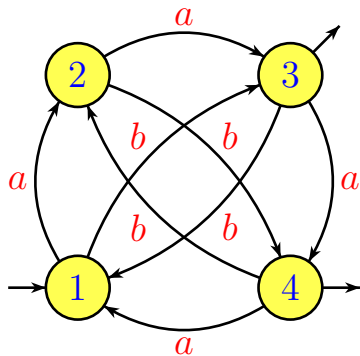


A group language



This group automaton recognizes the set of words u such that $\binom{u}{ab}$ is an **odd** number. Its transition monoid is a **2**-group of order **8**.

Another group automaton



Polynomial closure of the group languages

The **polynomial closure** of the group languages $\text{Pol } \mathcal{G}$ is the set of languages that are **finite unions** of languages of the form $L_0 a_1 L_1 \cdots a_k L_k$ where $k \geq 0$, each $a_i \in A$ and each L_i is a **group language**.

Theorem

*A regular language is **open** in the pro-group topology iff it belongs to $\text{Pol } \mathcal{G}$.*

Polynomial closure of the group languages

The **polynomial closure** of the group languages $\text{Pol } \mathcal{G}$ is the set of languages that are **finite unions** of languages of the form $L_0 a_1 L_1 \cdots a_k L_k$ where $k \geq 0$, each $a_i \in A$ and each L_i is a **group language**.

Theorem

*A regular language is **open** in the pro-group topology iff it belongs to $\text{Pol } \mathcal{G}$.*

Is it decidable?

Topology on the free group

Theorem (M. Hall Jr., 1950)

Every *finitely generated* subgroup of a finitely generated free group is *closed*.

Theorem (Ribes–Zalesskii, 1993)

Let H_1, \dots, H_n be f.g. subgroups of $FG(A)$. Then $H_1 \cdots H_n = \{h_1 \cdots h_n \mid h_1 \in H_1, \dots, h_n \in H_n\}$ is *closed*.

Several proofs, including one via **model theory**.

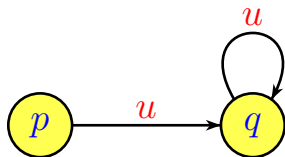


Closure of a regular language

Theorem (Pin 1991, Pin-Reutenauer 1991)

The closure of a *regular* language for d_G is still *regular* and can be computed effectively.

Intuitively, the only rule to apply is $\lim_{n \rightarrow \infty} x^{n!} = 1$.
In practice, just two rules, where u is any word.

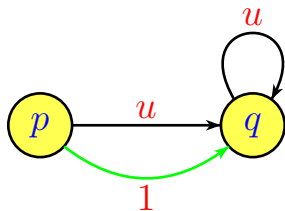


Closure of a regular language

Theorem (Pin 1991, Pin-Reutenauer 1991)

The closure of a *regular* language for d_G is still *regular* and can be computed effectively.

Intuitively, the only rule to apply is $\lim_{n \rightarrow \infty} x^{n!} = 1$.
In practice, just two rules, where u is any word.

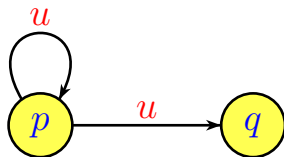
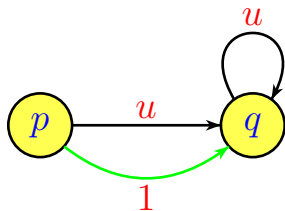


Closure of a regular language

Theorem (Pin 1991, Pin-Reutenauer 1991)

The closure of a *regular* language for d_G is still *regular* and can be computed effectively.

Intuitively, the only rule to apply is $\lim_{n \rightarrow \infty} x^{n!} = 1$.
In practice, just two rules, where u is any word.

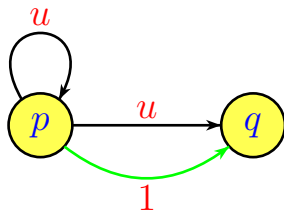
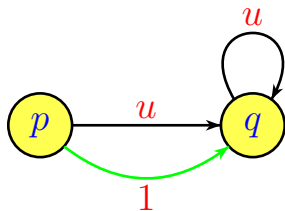


Closure of a regular language

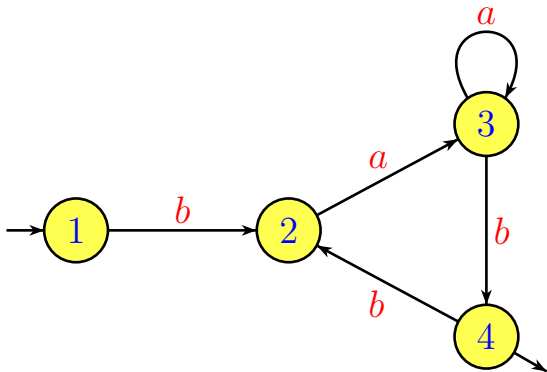
Theorem (Pin 1991, Pin-Reutenauer 1991)

The closure of a *regular* language for d_G is still *regular* and can be computed effectively.

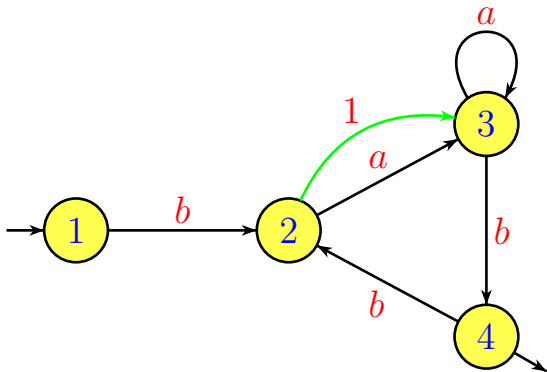
Intuitively, the only rule to apply is $\lim_{n \rightarrow \infty} x^{n!} = 1$.
In practice, just two rules, where u is any word.



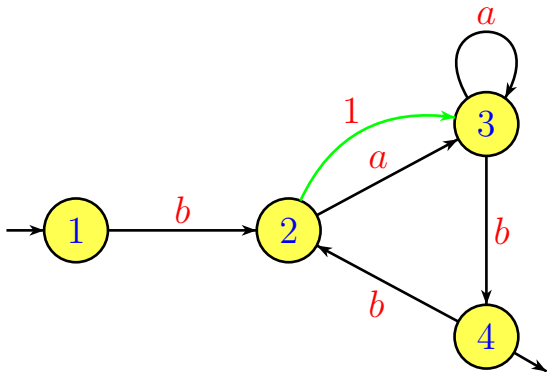
An example: $L = (ba^+b)^+$



An example: $L = (ba^+b)^+$

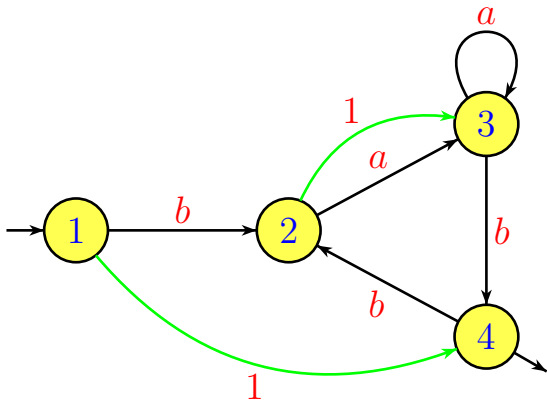


An example: $L = (ba^+b)^+$



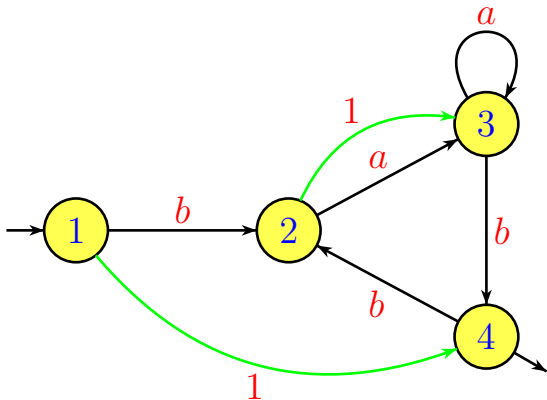
Now, there are a path from 1 to 4 and a loop around 4, both labeled by bb .

An example: $L = (ba^+b)^+$



Now, there are a path from 1 to 4 and a loop around 4, both labeled by bb .

An example: $L = (ba^+b)^+$



Now, there are a path from 1 to 4 and a loop around 4, both labeled by bb . The closure of L is $(ba^*b)^*$.

Part V

Some consequences

Some consequences in finite semigroup theory

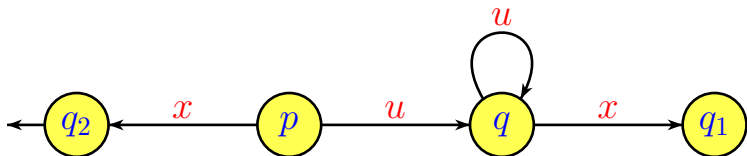
- The variety of finite monoids generated by the monoids of **partial injective maps** on a finite set is the variety of monoids whose **idempotent commute** [Margolis-Pin 87, Ash 87].

Some consequences in finite semigroup theory

- The variety of finite monoids generated by the monoids of **partial injective maps** on a finite set is the variety of monoids whose **idempotent commute** [Margolis-Pin 87, Ash 87].
- The **Mal'cev product** of a **decidable variety** of finite monoids with the variety of **finite groups** is still a **decidable** variety. (In general, Mal'cev products do not preserve decidability) [Ash 91].

Some consequences in automata theory

- One can **decide** whether a given regular language can be accepted by a **reversible automaton** [Pin 87] .
- A regular language is **open** in the pro-group topology iff its **minimal automaton** does not contain the following **pattern**, where u and x are words, and $q_1 \neq q_2$.



Part VI

Conclusion and open problems

Open problems

One can define **pro- p -group**, **pro-group**
pro-nilpotent, **pro-solvable** topologies, etc.

Algorithms are known to compute the closure of a regular language for the **pro-group**, the **pro- p** and the **pro-nilpotent** topology. It is still an **open problem** for the **pro-solvable** topology!

The problem amounts to deciding whether a partial group automaton can be **completed**, by **adding** an arbitrary number of states and transitions, into a **solvable group automaton**.






Beyond groups

One can build profinite topologies for other **varieties of finite monoids**. The free monoid A^* is then equipped with a structure of metric space, whose **completion** is a **compact** monoid.




These objects are still **mysterious** and are the key to the solution of numerous problems of the **theory of automata** and the topic of active research [Almeida, Auinger, Steinberg, Weil, etc.]






References I

-  C. J. Ash, Finite semigroups with commuting idempotents, in *J. Austral. Math. Soc. Ser. A* **43**, (1987), 81–90.
-  C. J. Ash, Inevitable Graphs: A proof of the type II conjecture and some related decision procedures, *Int. Jour. Alg. and Comp.* **1** (1991) 127–146.
-  J. Berstel, M. Crochemore et J.E. Pin, Thue-Morse sequence and p -adic topology for the free monoid, *Discrete Math.* **76** (1989) 89–94.




References II

-  M. Hall Jr, A topology for free groups and related groups, *Ann. Math.* **52** (1950) 127–139.
-  S. W. Margolis and J.E. Pin, Inverse semigroups and varieties of finite semigroups, *Journal of Algebra* **110** (1987), 306–323.
-  J.-E. Pin, Finite group topology and p -adic topology for free monoids, 12th ICALP, Lect. Notes Comp. Sci. **194** (1985) 445-455.





References III

-  J.-E. Pin, On the languages accepted by finite reversible automata, in *14th ICALP*, Berlin, 1987, Lect. Notes Comp. Sci. **267** (1987) 237–249.
-  J.-E. Pin, A topological approach to a conjecture of Rhodes, *Bulletin of the Australian Mathematical Society* **38** (1988), 421–431.
-  J.-E. Pin, Topologies for the free monoids, *Journal of Algebra* **137** (1991) 297–337.


References IV

-  J.-E. Pin, On reversible automata, in *Proceedings of the first LATIN conference*, São-Paulo, LNCS 583, (1992), 401–416.
-  J.-E. PIN, Topologie p -adique sur les mots, *Journal de théorie des nombres de Bordeaux* **5** (1993), 263–281.
-  J.-E. Pin, Polynomial closure of group languages and open sets of the Hall topology, *Theoret. Comput. Sci.* **169** (1996), 185–200.

References V

-  J.-E. Pin and C. Reutenauer, A conjecture on the Hall topology for the free group, *Notices of the London Math. Society* **23**, (1991), 356–362.
-  C. Reutenauer, Une topologie du monoïde libre, *Semigroup Forum* **18**, (1979), 33–49.
-  C. Reutenauer, Sur mon article “Une topologie du monoïde libre”, *Semigroup Forum* **22** (1981) 93–95.
-  L. Ribes and P.A. Zalesskii, On the profinite topology on a free group, *Bull. London Math. Soc.* **25** (1993) 37–43.

References VI

-  L. Ribes and P.A. Zalesskii, The pro- p topology of a free group and algorithmic problems in semigroups, *Int. Jour. Alg. and Comp.* **4** (1994) 359–374.