Random Automata

Parisa Babaali

Applications of Automata

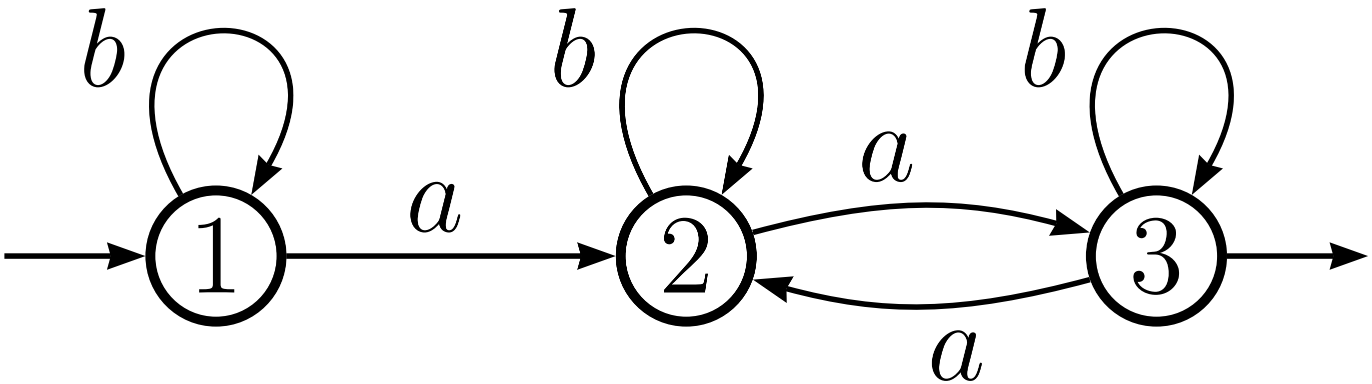
- Bioinformatics
- Compilers
- Computer-aided verification
- DNA/molecular/membrane computing
- Pattern-matching
- Speech and speaker recognition
- Digital libraries
- Natural Language Processing
- Quantum computing

Deterministic Finite Automata

Definition 1 *A Deterministic Finite Automaton(DFA) is a 5-tuple* $(Q, \Sigma, \delta, q_0, F)$*, where*

- Q *is a finite set of states,*
- Σ *is a finite input alphabet,*
- $q_0 \in Q$ *is the initial state,*
- $F \subset Q$ *is the set of final states,*
- δ *is the transition function mapping* $Q \times \Sigma$ *to* Q*.*

Example of a DFA



A DFA with 3 state, $\delta(1, a) = 2$ and $\delta(1, b) = 1$. Extend δ by defining: $\delta(1, ab) = 2$ and $\delta(2, bba) = 3$.

Regular Languages

$$\Sigma^* = \{w | w \text{ is a word in the alphabet } \Sigma\}$$

example a, ab, ba, babaa, bbbbbaaaaa are words in Σ^*.

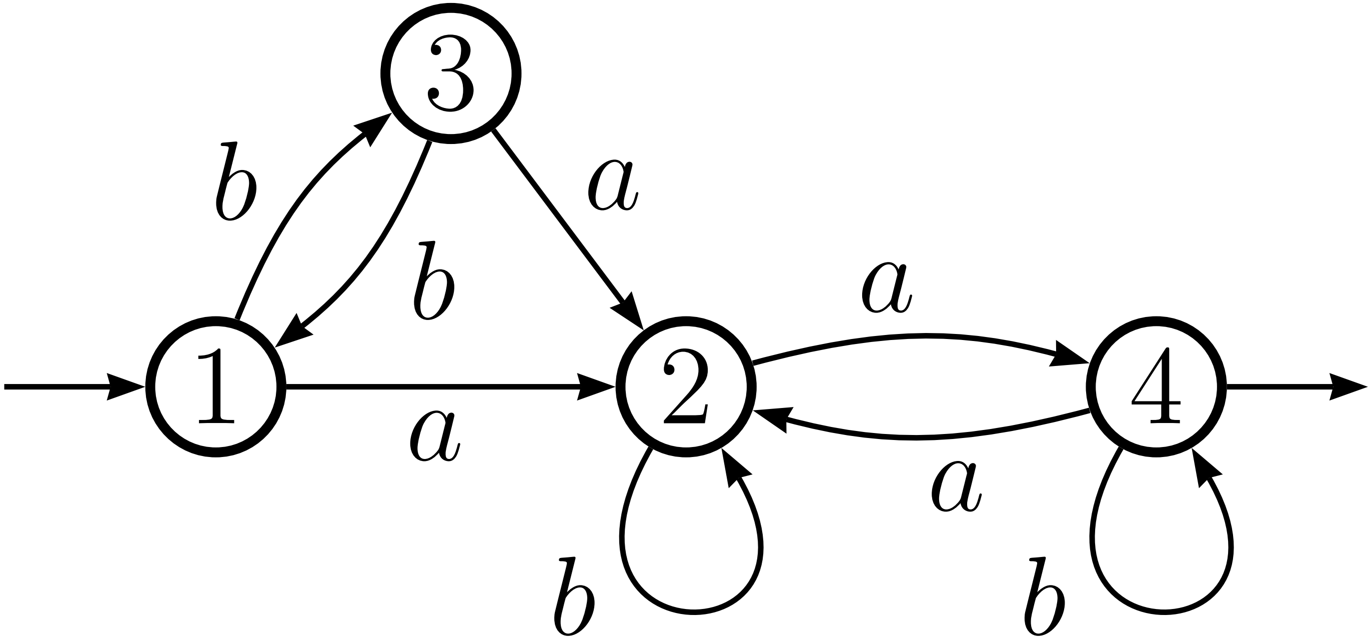
$x \in \Sigma^*$ is accepted by a DFA M if $\delta(q_0, x) = q$ form some $q \in F$. The language accepted by M is

$$L(M) = \{x \in \Sigma^* | \delta(q_0, x) \in F\}$$

A language is *regular* if it is the set accepted by some finite automaton.

Minimal Automata

There are many automata accepting a given regular language.



Minimal Automata

Theorem 1 *(Nerode-Myhill) There is an automata with the smallest number of states accepting a regular language* L *and it is unique.*

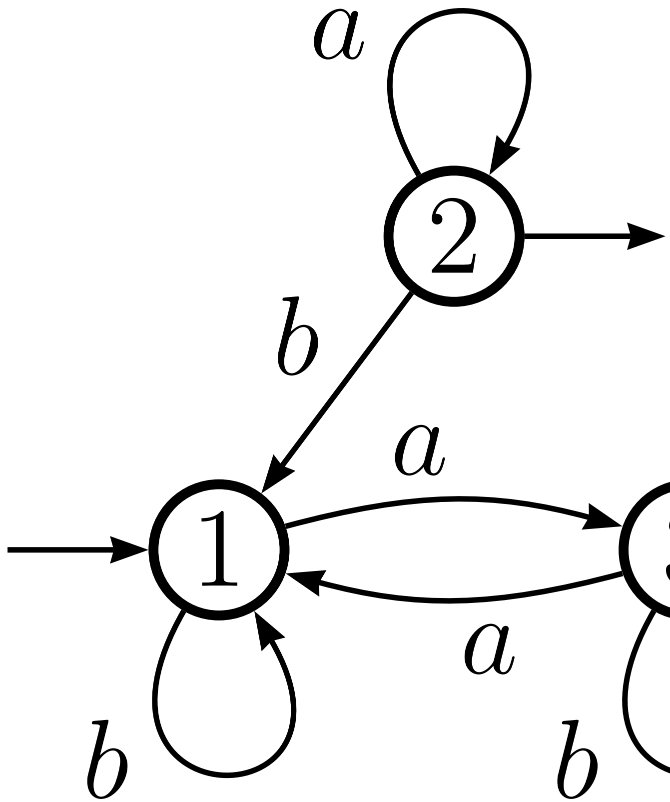
There are algorithms to find the minimal automaton of a given regular language, the coplexity, however has order n^2 where n is the number of states of an automaton.

Random Accessible Automata

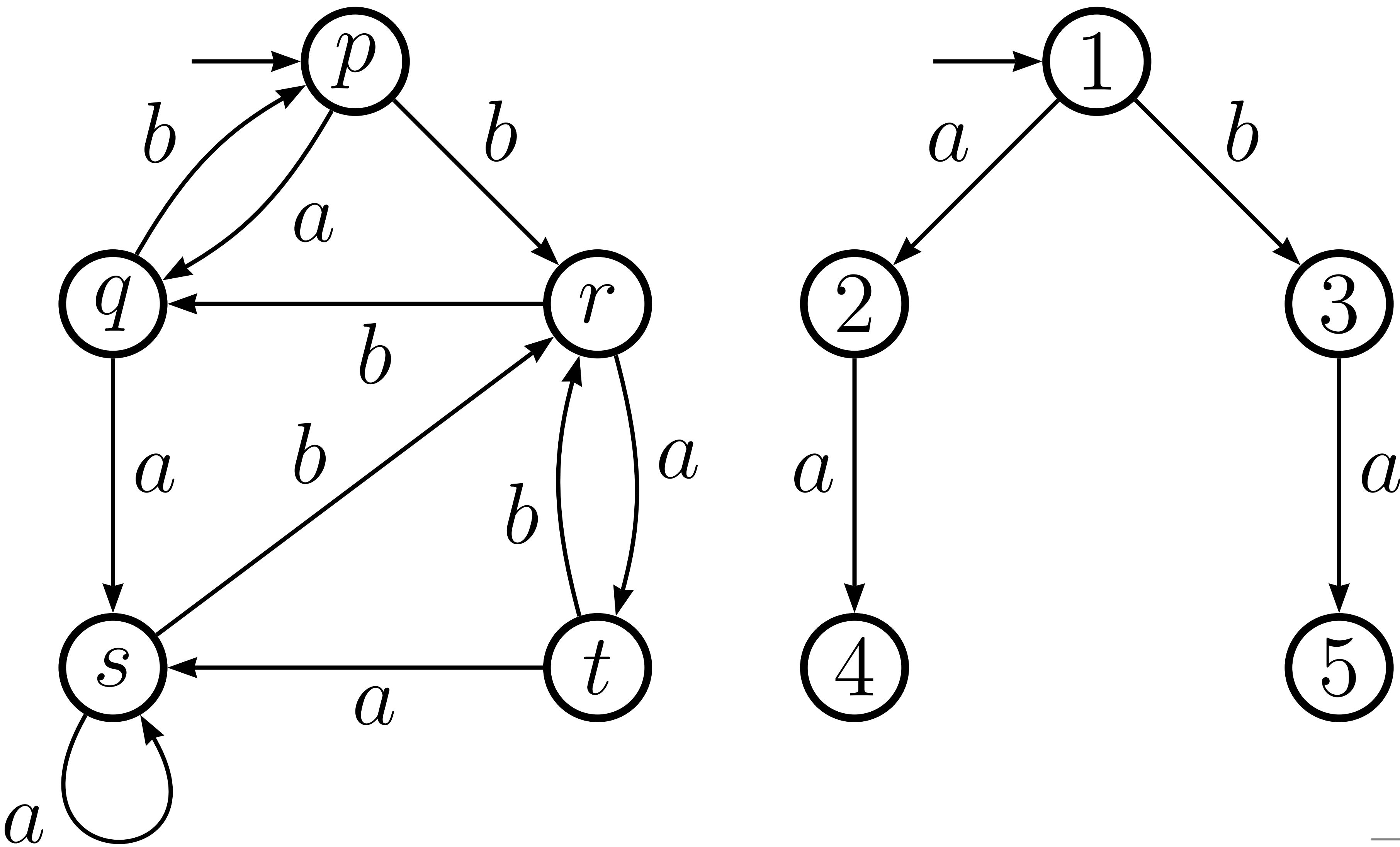
Definition 2 *An automaton is said to be* accessible *if for every state $q \in Q$ there exist a word w with $\delta(q_0, w) = q$.*

Observation Every accessible automata has a maximal spanning subtree. We are interested in the tree obtained by a breadth first traverse.

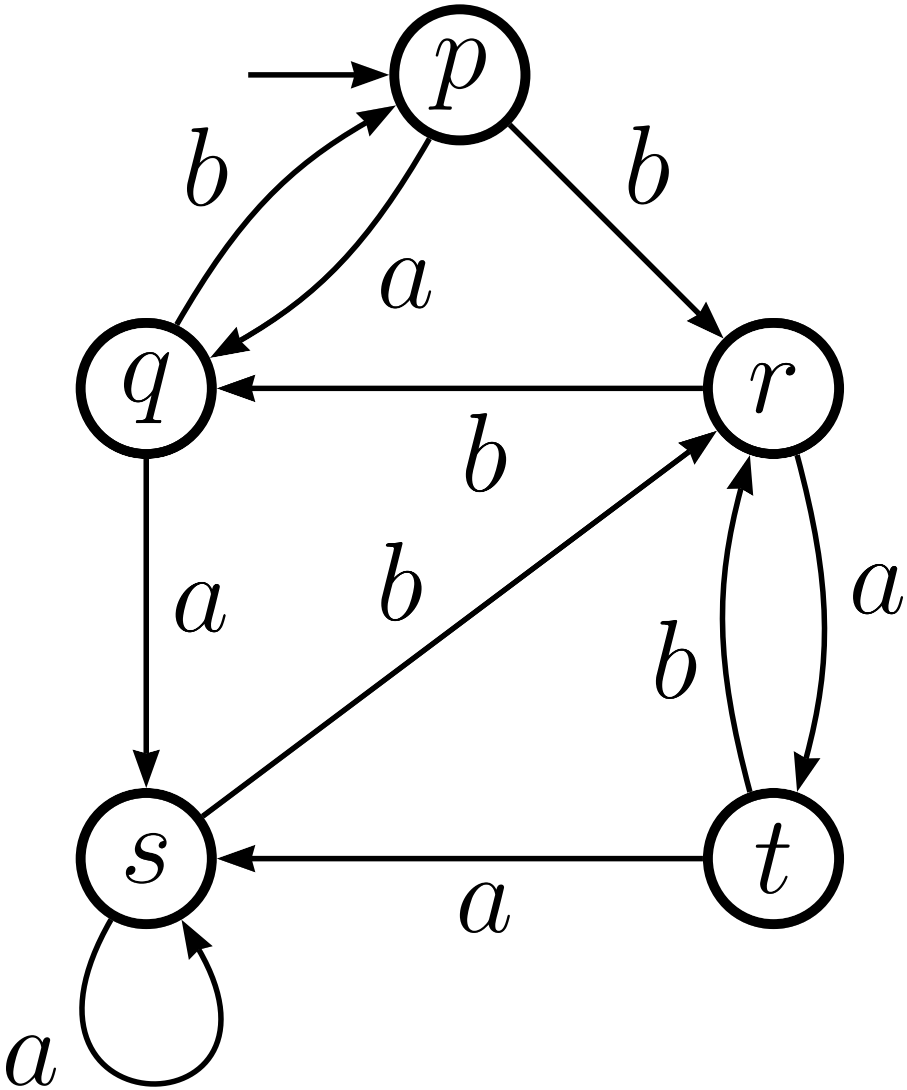
Example

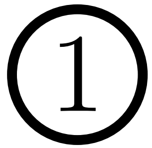


Example

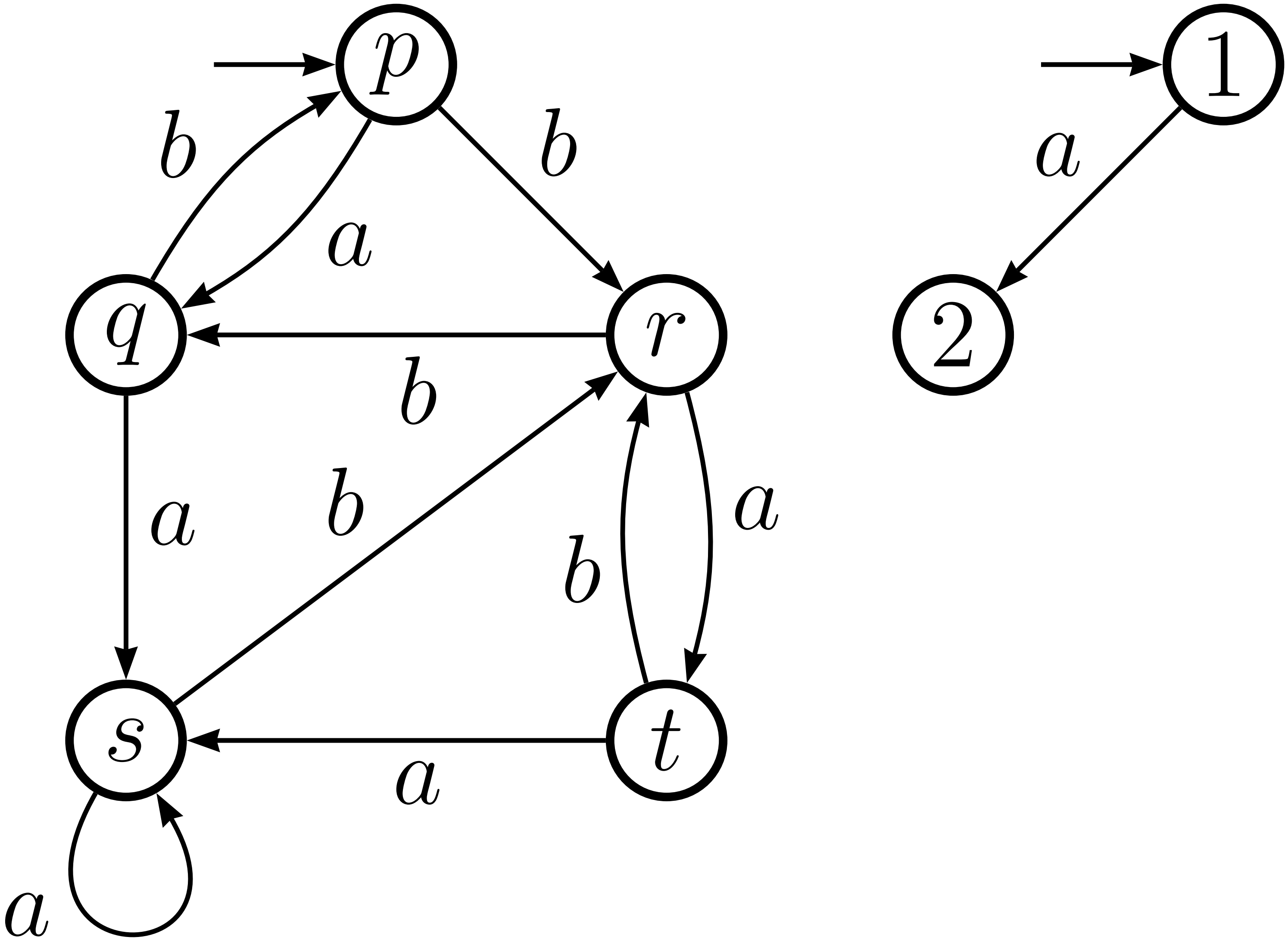


1

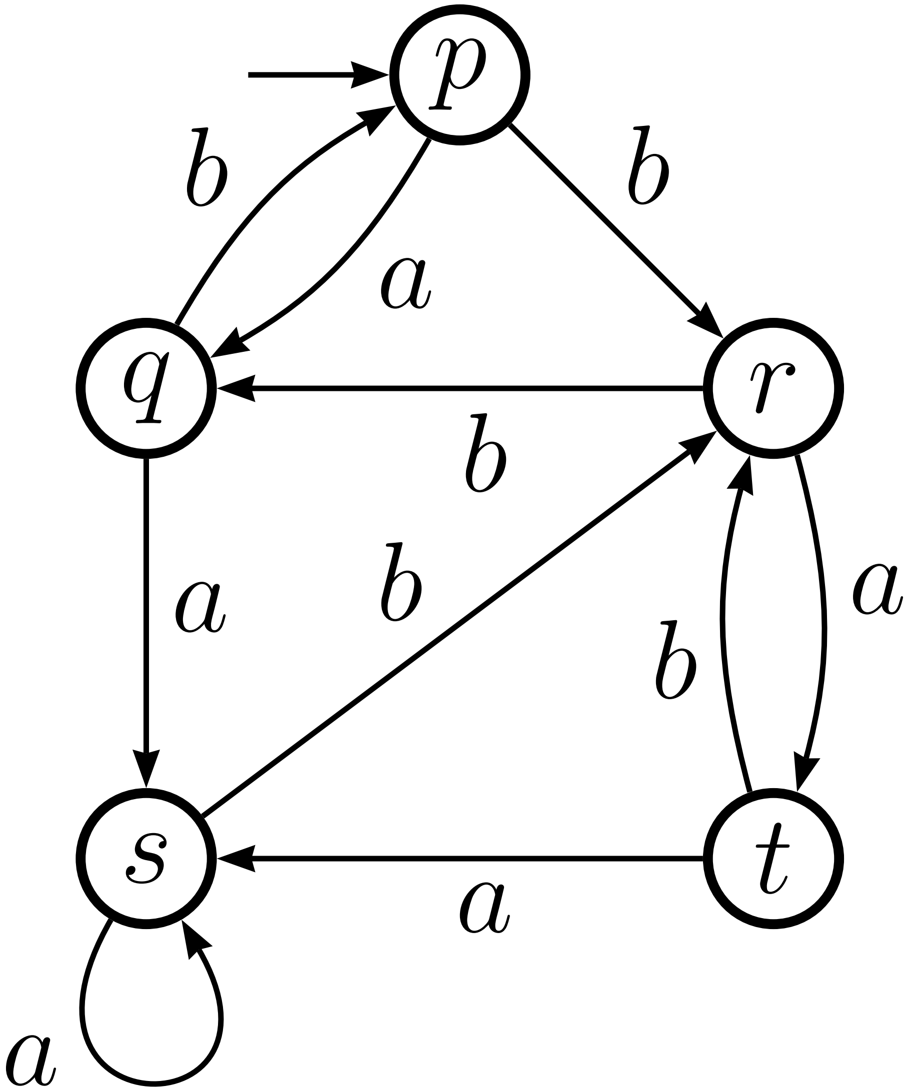


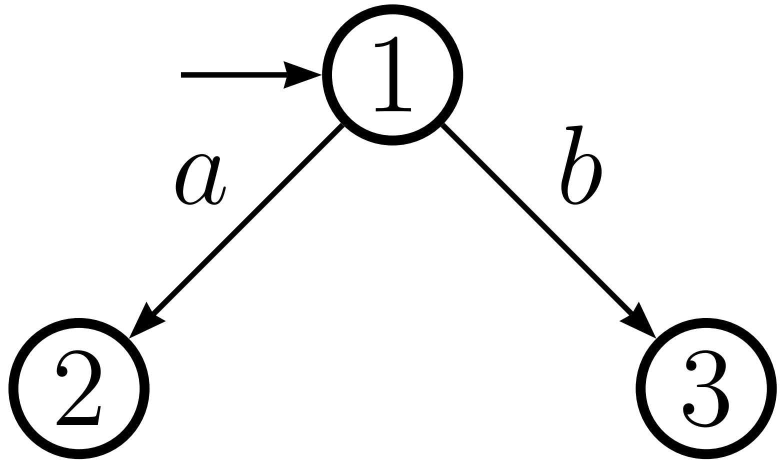


1 1

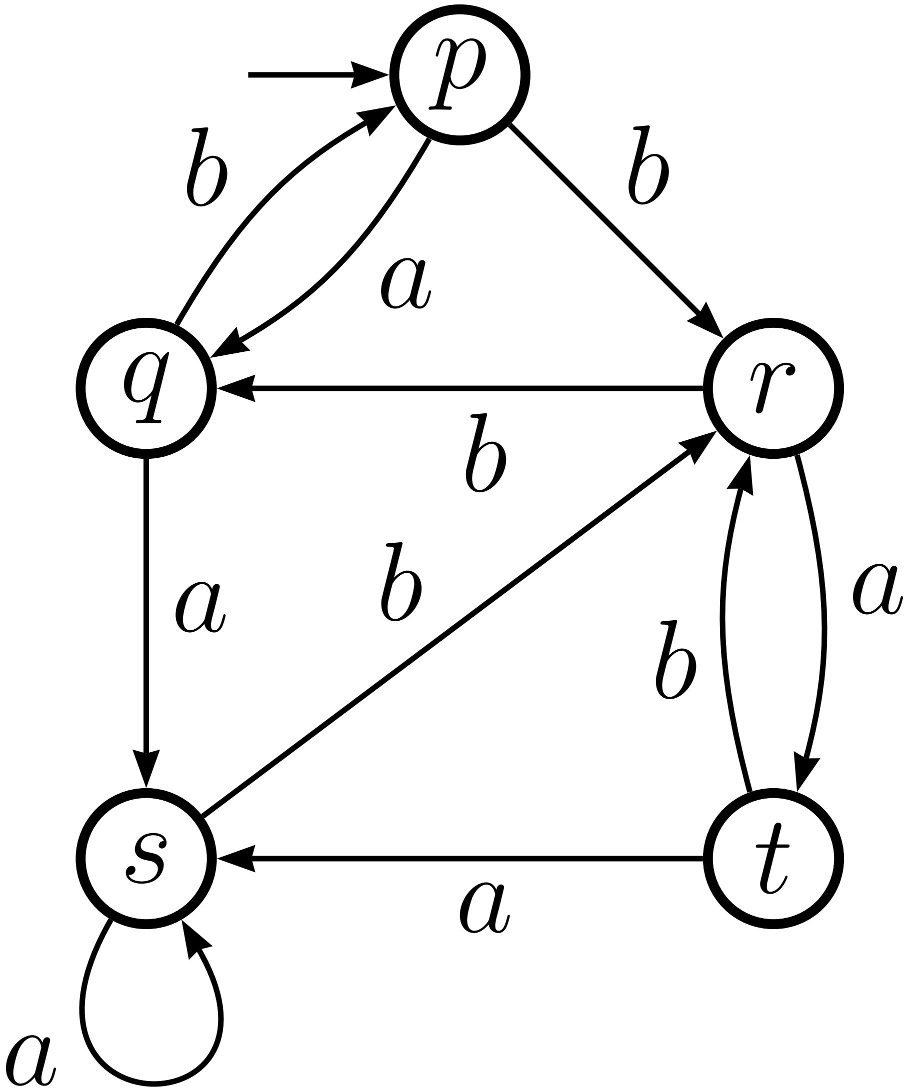


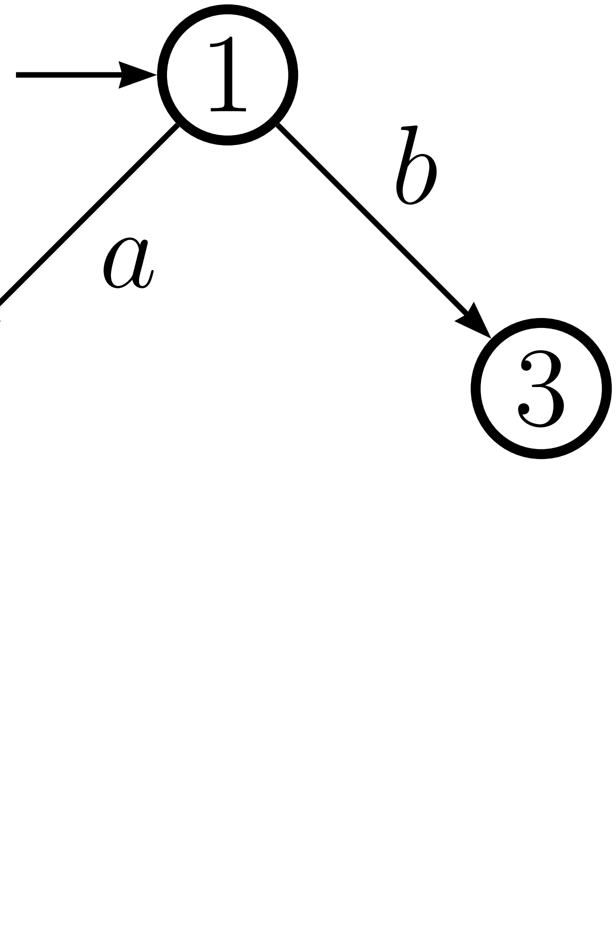
1 11



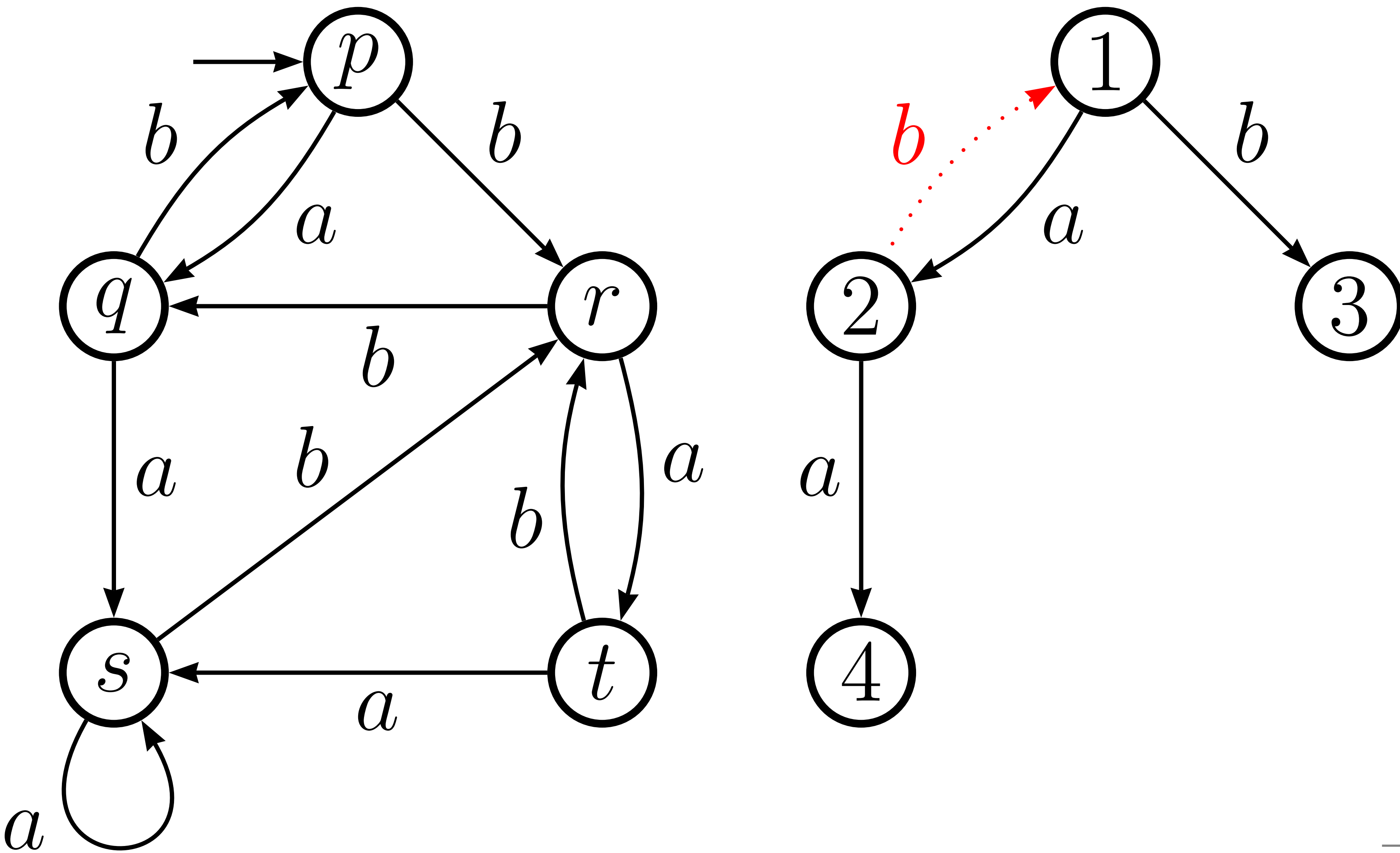


1 111 1

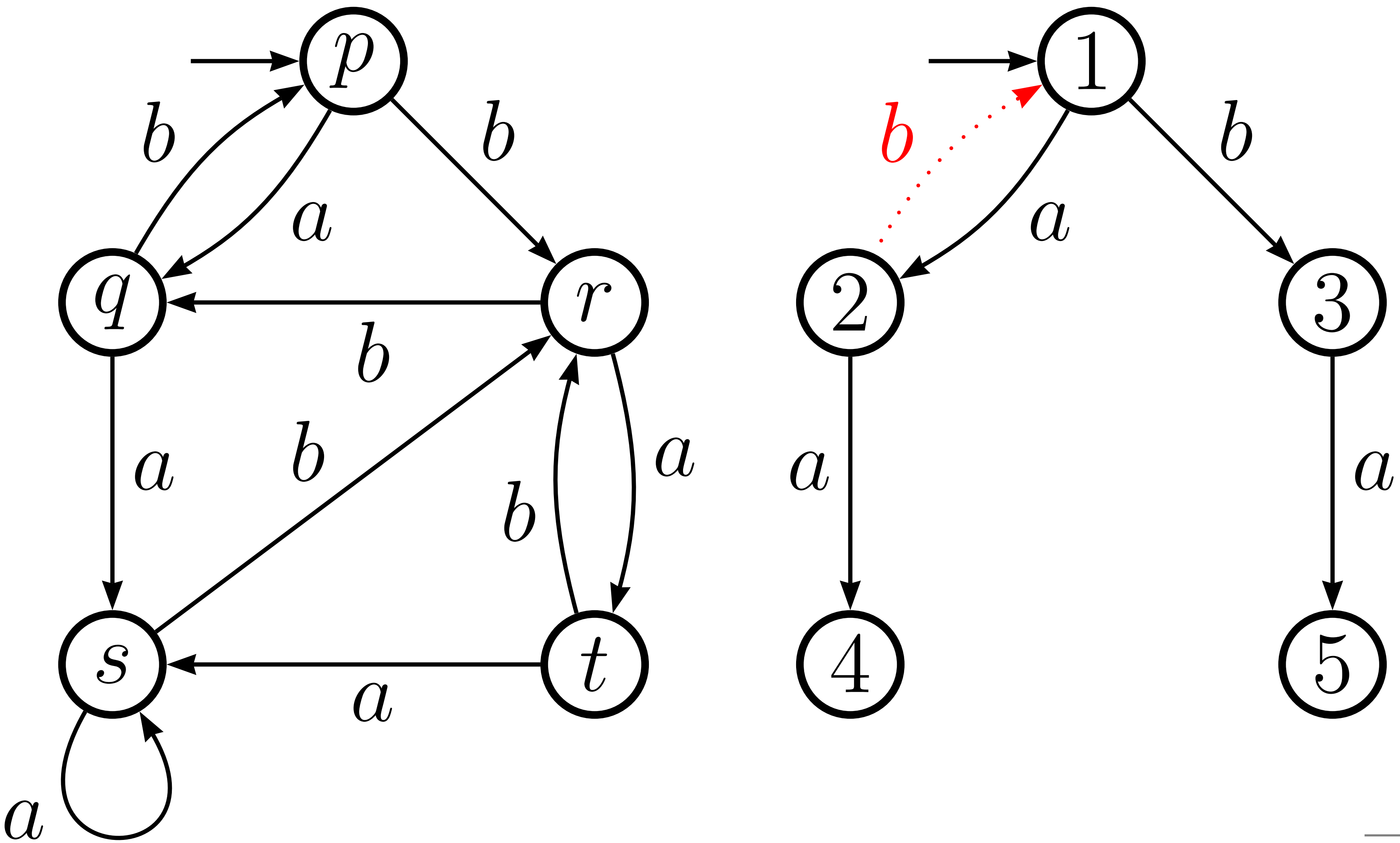




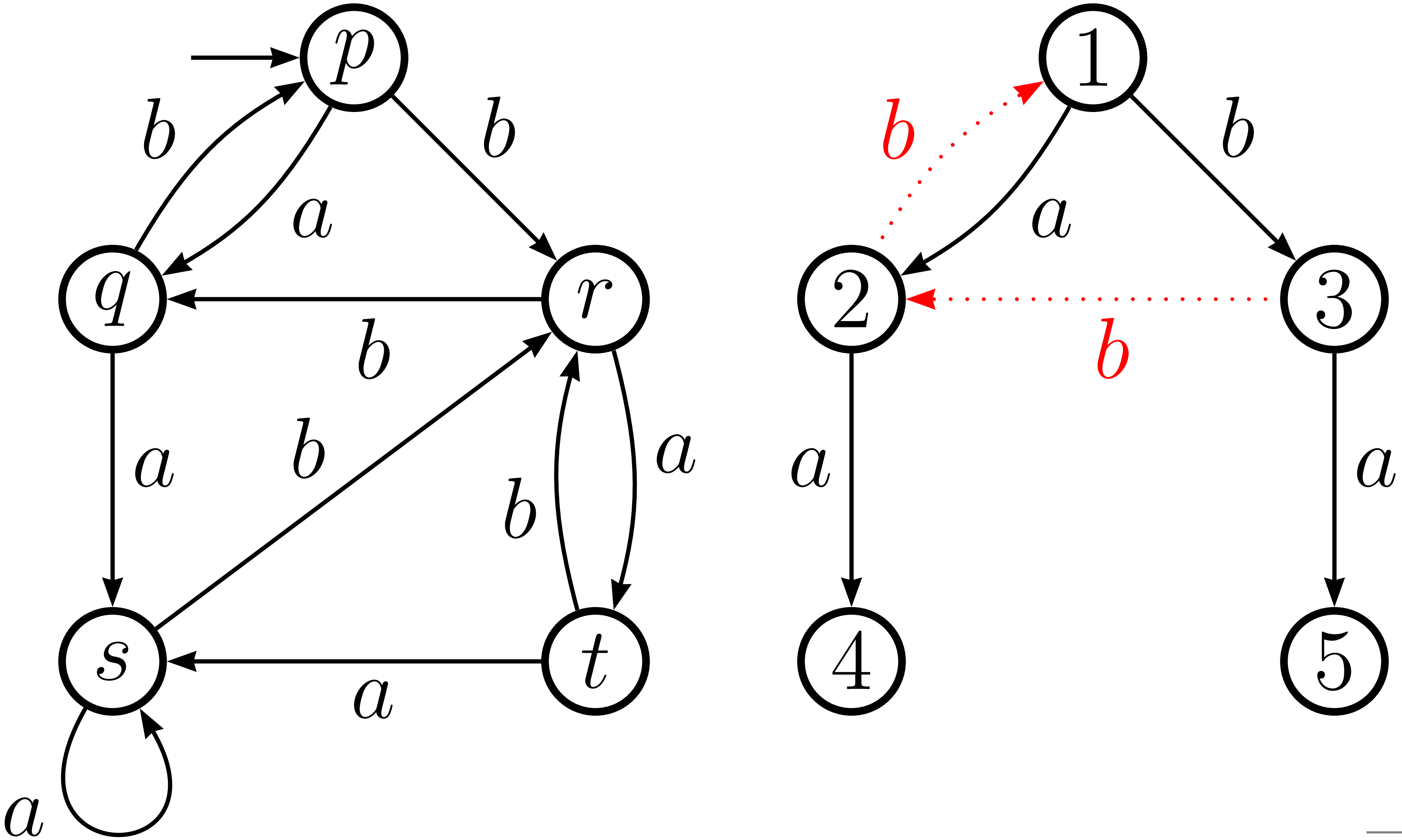
1 11 10



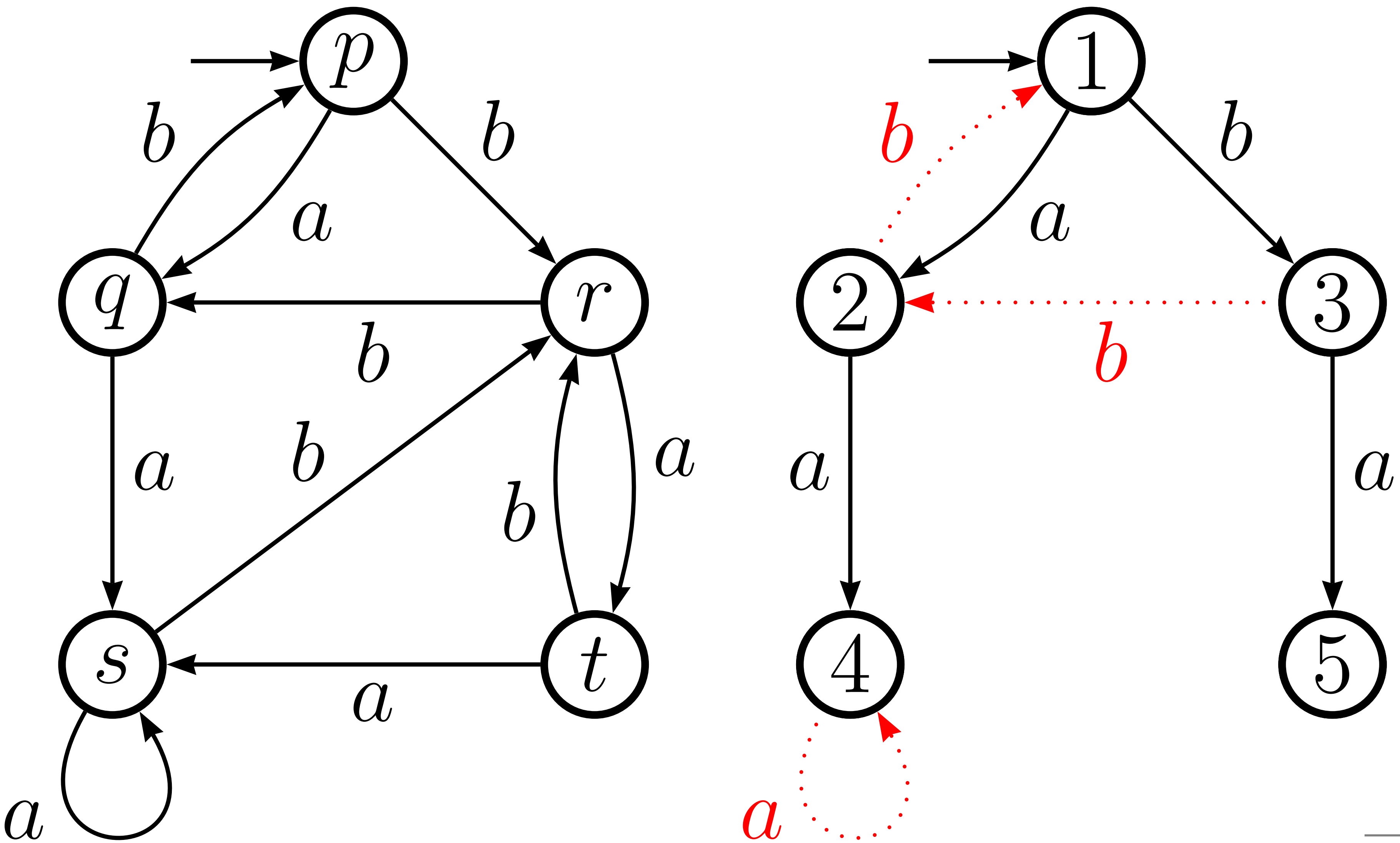
1 11 10 1



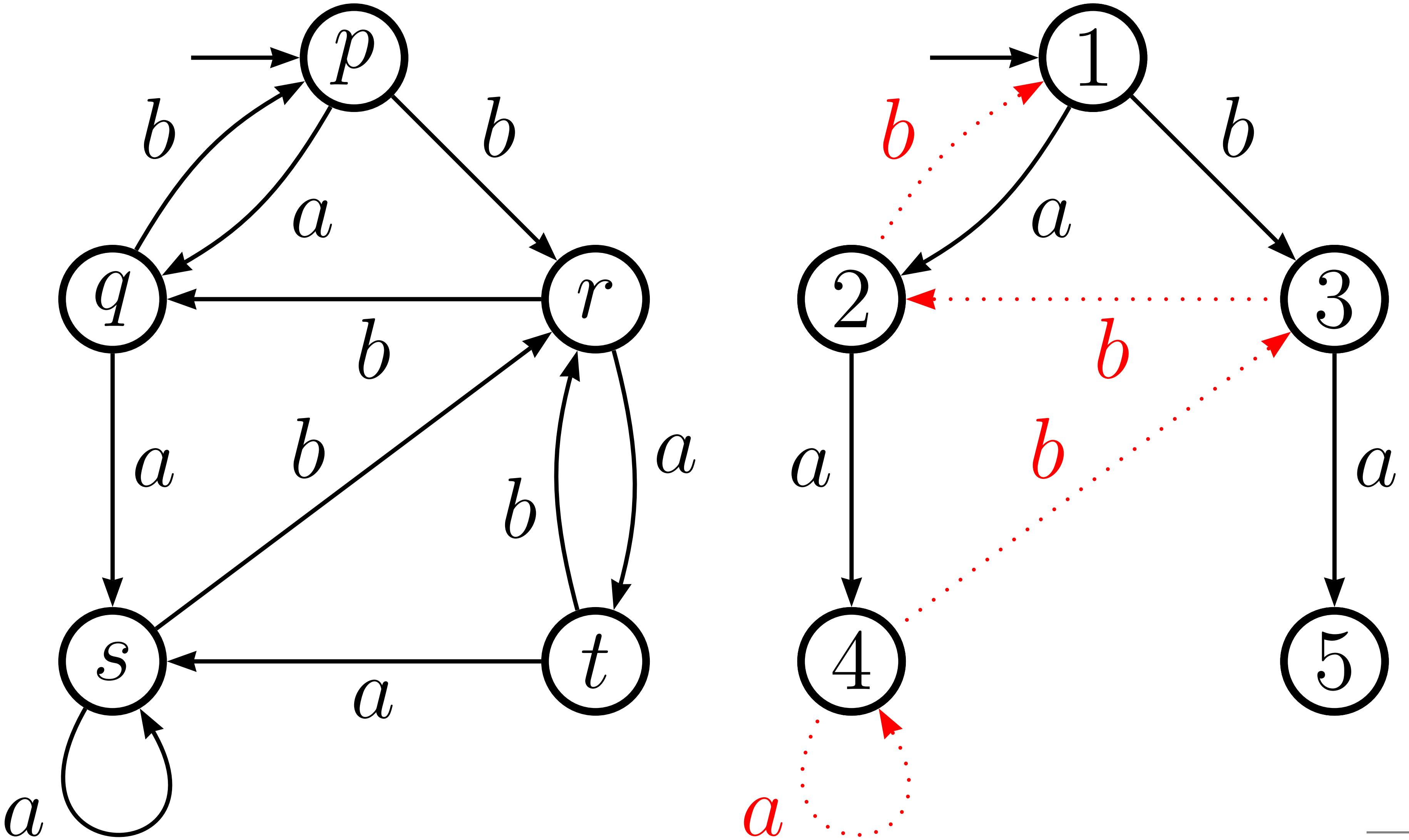
1 11 10 10



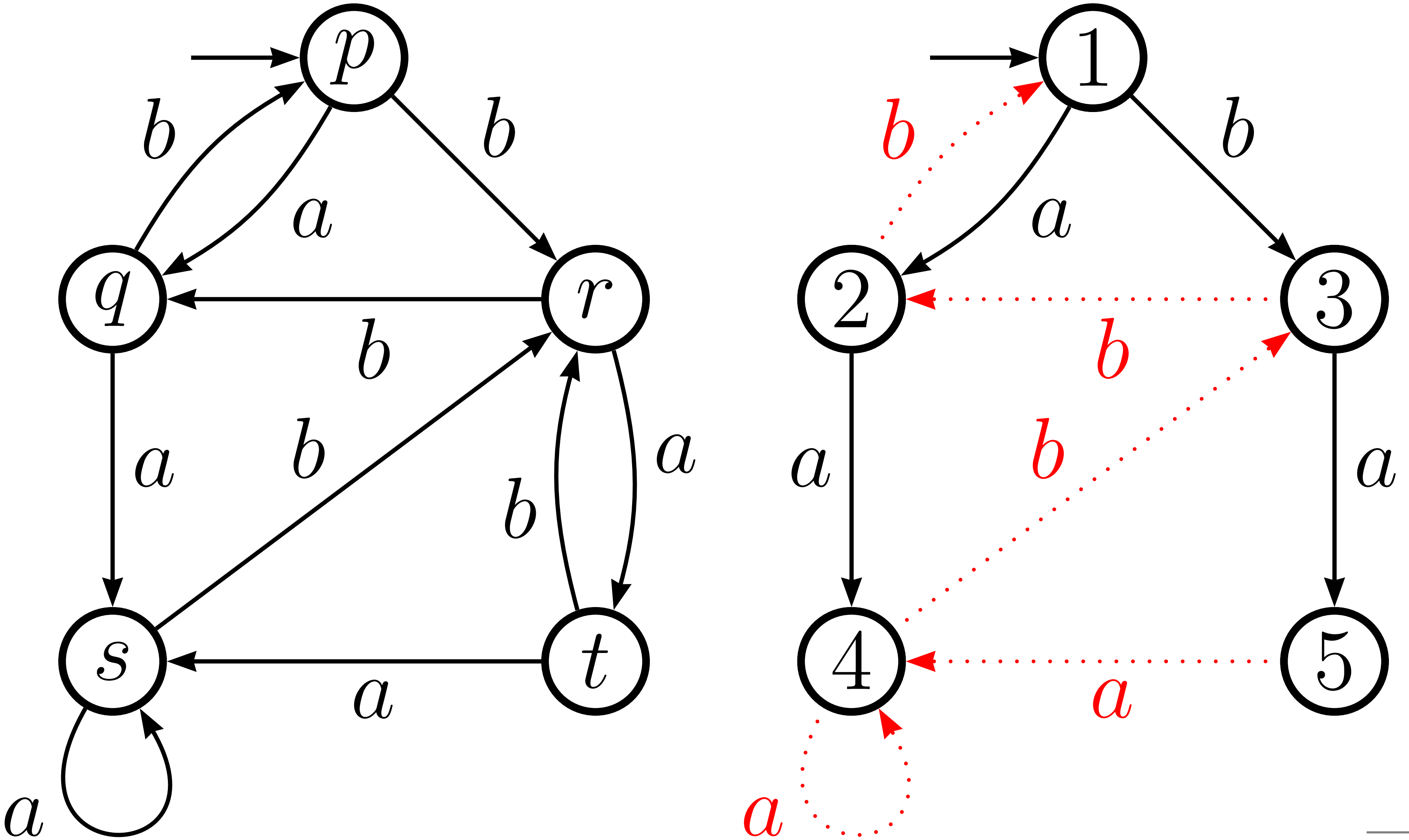
1 11 10 10 0



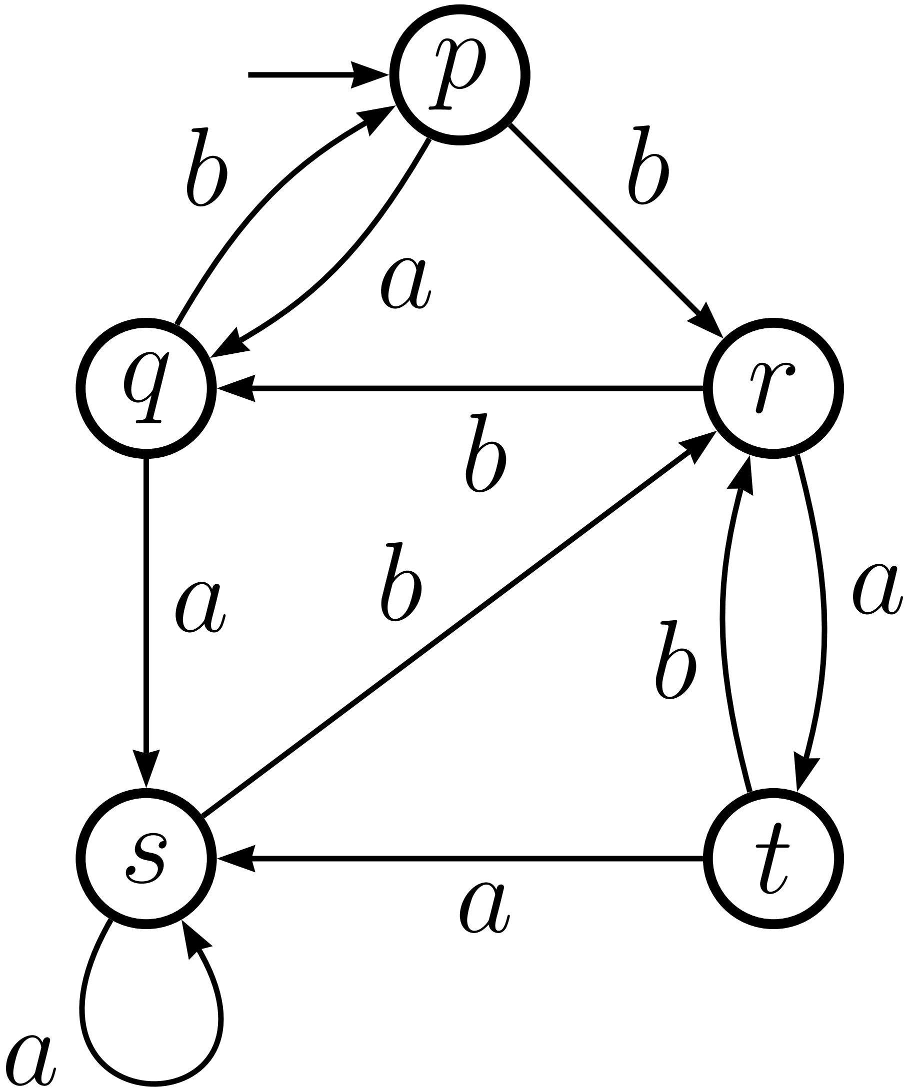
1 11 10 10 00

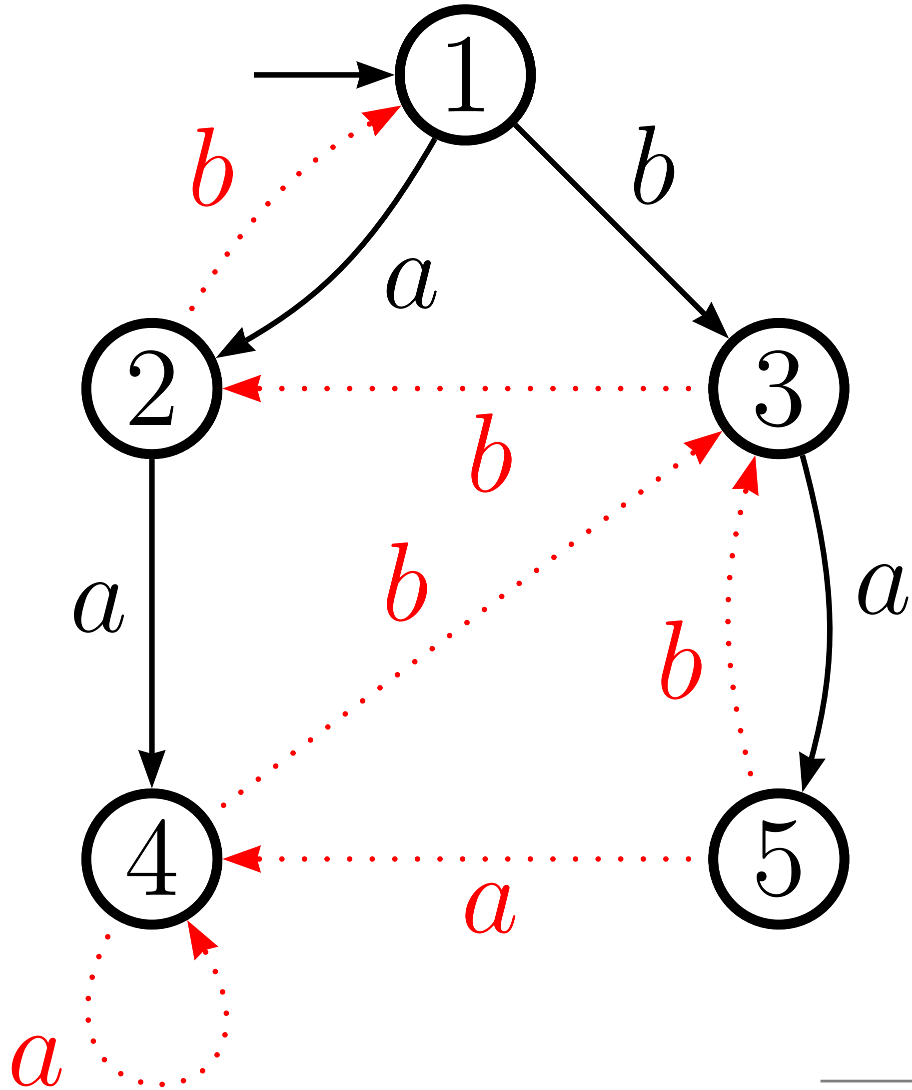


1 11 10 10 00 0



1 11 10 10 00 00





Transition Structure

A deterministic transition structure DTS is a 4-tuple (Q, Σ, δ, q_0). A DTS is a DFA without final states.

Call $\mathfrak{A}$ the set of all DTS's on the alphabet Σ.

For every DTS with n states there are 2^n possible automata.

Probability Measures on $\mathfrak{T}$

To a binary labeled tree $T \in \mathfrak{T}$ with n vertices assign a probability $\mu_p(T) = \mathbb{P}(T) = p^{n-1}(1-p)^{n+1}$.

claim $(\mathfrak{T}, \mu_p)$ is a probability space.

Proof There are $C_n = \frac{1}{n+1}\binom{2n}{n}$ trees with n vertices and

$$\sum_{n=1}^{\infty} C_n p^{n-1}(1-p)^{n+1} = 1.$$

Measures on Automata

Theorem 2 *Let* T *be a tree with* $\pi(T) = (i_1, i_2, \cdots, i_{2n})$*,let* n_T *be the number of transition structures having* T *as a spanning subtree and let* $x_k = 1 + \sum_{j=1}^{k} i_j$ *then:*

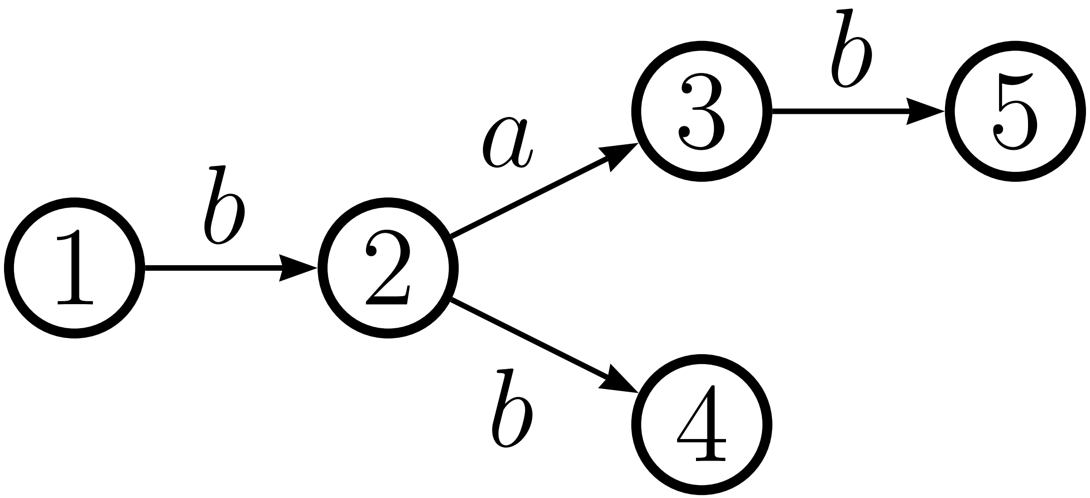
$$n_T = \prod_{k=1}^{2n} {x_k}^{1-i_k}$$

Now for an automata $\mathcal{A}$ with spanning tree T let:

$$\mu(\mathcal{A}) = \frac{1}{2^n n_T} \mu_p(T)$$

$(\mathfrak{A}, \mu)$ is a probability space.

Example



This tree has :

$$\pi(T) = (1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0)$$

There are $2^5 \times 1 \times 4 \times 5 \times 5 \times 5 \times 5 = 2^5 \times 4 \times 5^4$ automata having T as a subtree.

The Mean Number of Vertices of μ_p

Definition 3 *The mean size of elements of $\mathfrak{A}$ with respect to μ_p is:*

$$M_{\mu_p} = \sum_{T \in \mathfrak{T}} s(T)\mu_p(T)$$

where $s(T)$ is the number of nodes of T.

we can show:

$$M_{\mu_p} = \frac{1}{1 - 2p}$$

$M_{\mu_p} \to \infty$ as $p \to \frac{1}{2}$.

Size functions

Definition 4 *A size function on* A *is a map* $s : A \longrightarrow \mathbb{N}$ *such that for all* $x \in A$*;* $s^{-1}(x)$ *is a finite set.*

Definition 5 *A stratification of* A *is an increasing sequence of finite subsets of* A *with*

$$A_0 \subseteq A_1 \subseteq A_2 \cdots$$

where each A_i *is finite and*

$$\bigcup_{i=1}^{\infty} A_i = A$$

It's clear that given a size function s on A, one can find a stratification of A by $A_0 = s^{-1}(\{0\})$ and $A_1 = s^{-1}(\{0, 1\})$, ...

Asymptotic Density

Let A be a set equipped with a size function and A_n be the set of all elements of size n.

Definition 6 *The spherical asymptotic density of a set* $M \subseteq A$ *is defined to be the limsup:*

$$\rho(M) = \limsup_{n \to \infty} \frac{|A_n \cap M|}{|A_n|}$$

A set M is generic if $\rho(M) = 1$.

Let $s : \mathfrak{A} \to \mathbb{N}$ with $s(\mathcal{A}) =$ number of states of $\mathcal{A}$ and let $a_n = |\mathfrak{A}_n|$

Asymptotic Density

Question Do we know the number of deterministic automata with n states a_n ?

- Bassino and Nicaud (February 07) have shown that the asymptotic behavior of a_n is almost like the Sterling number of the second kind $\left\{ \begin{matrix} 2n \\ n \end{matrix} \right\}$.

Results

Let $\mathfrak{M}$ be the set of all minimal automata, and let $\mathfrak{E}$ be the set of all automata whose language has exponential growth.

Theorem 3 *Conditioned on* $\mathfrak{M}$ *,* $\mathfrak{E}$ *is a generic set.*

$$\rho(\mathfrak{E}|\mathfrak{M}) = 1$$

In other word most regular languages have exponential growth.

Minimal Automata

Question: What is the asymptotic density of $\mathfrak{M}$?

- Nicaud (Experiment Results, 2000) :

$$\rho(\mathfrak{M}) \approx 0.8$$

- Champarnaud, Paranthoen (Experimental Result, 2005):
  For alphabet of size more than 2 :

$$\rho(\mathfrak{M}) \approx 1$$

Asymptotic Density of Minimal Automata

Theorem 4 *The set of all non-minimal automata has positive asymptotic density, it has the lower bound*

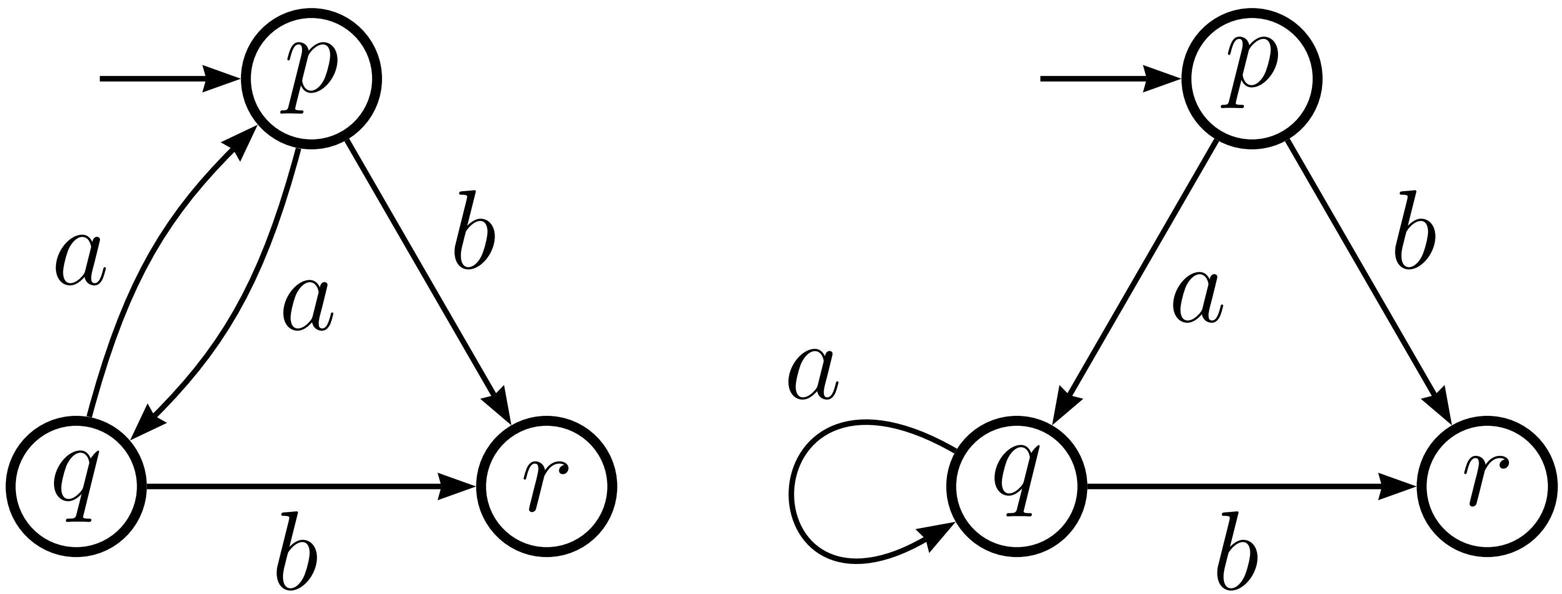
$$\rho(\mathfrak{M}^c) \geq 0.1$$

Asymptotic Density of Minimal Automata

idea of proof

States $p, q \in Q$ are equivalent if for all $x \in \Sigma^*$;

$$\delta(p, x) \in F \iff \delta(q, x) \in F$$

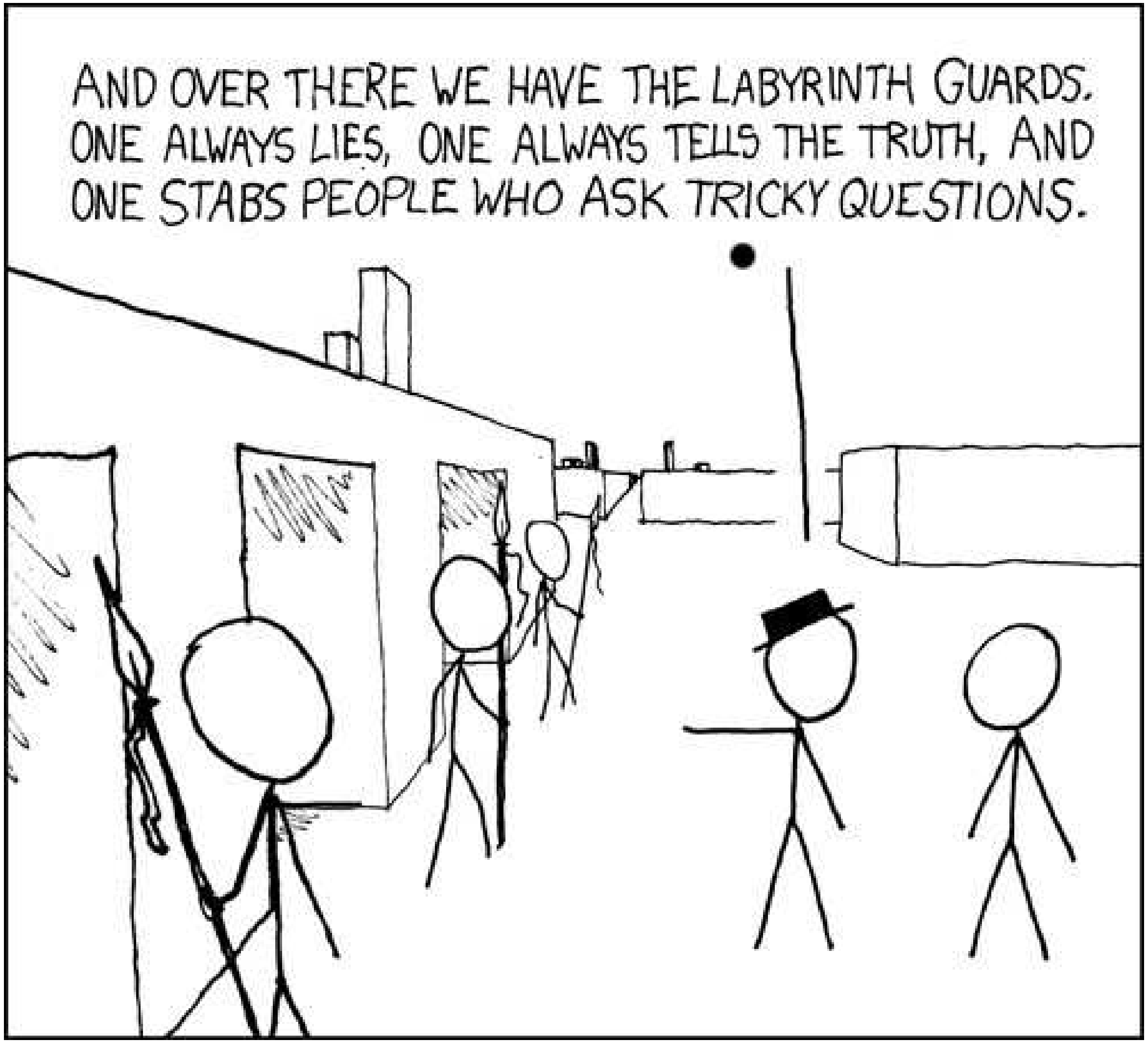


Asymptotic Density of Minimal Automata

p and q are equivalent states if and only if either both are final states or both are not.

- $\mathfrak{A}_T$ The set of all automata having T as a BFS subtree.
- C_T Those which are not minimal
- $\frac{|C_T|}{|\mathfrak{F}_T|} = \frac{2}{9}$.
- $$\rho(C) = \lim_{n\to\infty} \frac{\frac{2}{9}C_{n-2}}{C_n} = \frac{2}{9}\frac{C_{n-2}}{C_n} = \frac{2}{9}\frac{1}{4^2} = \frac{1}{72}$$

Questions ?


AND OVER THERE WE HAVE THE LABYRINTH GUARDS. ONE ALWAYS LIES, ONE ALWAYS TELLS THE TRUTH, AND ONE STABS PEOPLE WHO ASK TRICKY QUESTIONS.