On rational graphs

Christophe Morvan

On rational graphs

Christophe Morvan

IRISA, Campus de Beaulieu, 35042 Rennes, France
christophe.morvan@irisa.fr

Abstract. Using rationality, like in language theory, we define a family of infinite graphs. This family is a strict extension of the context-free graphs of Muller and Schupp, the equational graphs of Courcelle and the prefix recognizable graphs of Caucal. We give basic properties, as well as an internal and an external characterization of these graphs. We also show that their traces form an AFL of recursive languages, containing the context-free languages.

1 Introduction

When dealing with computers, infinite graphs are natural objects. They emerge naturally in recursive program schemes or communicating automata, for example. Studying them as families of objects is comparatively recent: Muller and Schupp (in [MS 85]) first captured the structure of the graphs of pushdown automata, then Courcelle (in [Co 90]) defined the set of regular (equational) graphs. More recently Caucal introduced (in [Ca 96]) a characterization of graphs in terms of inverse (rational) substitution from the complete binary tree. Step by step, like Chomsky's languages family, a hierarchy of graph families is built: the *graphs of pushdown automata*, *regular graphs* and *prefix-recognizable graphs*.

To define infinite objects conveniently, we have to use finite systems. For infinite graphs, two kinds of finite systems are employed: internal systems or external systems. Roughly speaking an internal characterization is a *machine* producing the arcs of the graph. An external characterization yields the structure of the graph (usually "up to isomorphism"). There is, of course a relationship between internal and external characterization: for example the pushdown automata are an internal characterization of the connected regular graphs of finite degree whereas the *deterministic graph grammars* are an external system for the family of regular graphs.

The purpose of this article is to give both internal and external characterization of a wider family of graphs. Using words for vertices, *rationality* (like in language theory) will provide an internal characterization; it will also give basic results for this family: for example rational graphs will be recognized by transducers; a rational graph is a recursive set; determinism for rational graphs will be decidable. Then *inverse substitution* from the complete binary tree (like in [Ca 96]) will be an external characterization of this family. Strangely this extension will prove to be a slight extension of the *prefix-recognizable graphs*: instead of taking the inverse image of the complete binary tree by a rational substitution

we will consider the inverse image of the complete binary tree by a linear substitution (*i.e.,* a substitution where the image of each letter is a linear language). Finally properties of the traces of these graphs will be investigated: we will show that the traces of these graphs form an *abstract family of (recursive) languages* containing the context-free languages.

2 Rational graphs

In this section we will define a new family of infinite graphs, namely the set of *rational graphs*. We will state some results for this family and give examples of rational graphs.

2.1 Partial semigroups

This paragraph introduces rationality for partial semigroups and uses this notion to give a natural introduction for rational graphs.
We start by recalling some standards notations: for any set E, its cardinal is denoted by $|E|$; its powerset is denoted by 2^E. Let the set of nonnegative integers be denoted by $\mathbb{N}$. A semigroup S is a set equipped with an operation $\cdot : S \times S \to S$ such that: for all u, v in S there exists w in S such that $\cdot(u, v) = w$ denoted by $u \cdot v = w$ and this operation is associative (*i.e.,* $\forall u, v, w \in S, (u \cdot v) \cdot w = u \cdot (v \cdot w)$. Finally, a monoid M is a semigroup with a (unique) neutral element (denoted ε along these lines) *i.e.,* an element $\varepsilon \in M$ such that for all element u in M $u \cdot \varepsilon = \varepsilon \cdot u = u$.
Now, a *partial semigroup* is a set S equipped with $\cdot : S \times S \to S$, a partial operation, with $\mathcal{D} \subseteq S \times S$ the domain of $\cdot$; set $\mathcal{D}$ need not be $S \times S$. Moreover we impose this operation to be *associative* as follows: $[(u, v) \in \mathcal{D} \wedge ((u \cdot v), w) \in \mathcal{D}] \Leftrightarrow [(v, w) \in \mathcal{D} \wedge (u, (v \cdot w)) \in \mathcal{D}]$ and in that case, $u \cdot (v \cdot w) = (u \cdot v) \cdot w$. Meaning that if multiplication is defined on the one side, then it is defined on the other side and both agree.
Notice that a partial semigroup S such that $\mathcal{D}$ is $S \times S$ is a semigroup.

Example 2.1. Given two semigroups $(S_1, \cdot_1)$ and $(S_2, \cdot_2)$ such that $S_1 \cap S_2$ is empty. The union $S = S_1 \cup S_2$, with the partial operation $\cdot$ defined as $\cdot_1$ over the elements of S_1 and $\cdot_2$ over the element of S_2, is a partial semigroup.

Taking a new element $\perp$ we complete any partial semigroup S into a semigroup $S \cup \{\perp\}$ by extending its operation $\cdot$ as follows:
$a \cdot b = \perp$ for all $a, b \in S \cup \{\perp\}$ such that $(a, b) \notin \mathcal{D}$.
Also the product $S \times S'$ of two partial semigroups S and S' is a partial semigroup for operation $\cdot$ defined componentwise:
$(a, a') \cdot (b, b') = (a \cdot b, a' \cdot b')$ for all $(a, b) \in \mathcal{D}$ and $(a', b') \in \mathcal{D}'$.
In order to define the rational subsets of a partial semigroup, we have to extend its operation to its subsets:

$$A \cdot B := \{a \cdot b \mid a \in A \wedge b \in B\} \text{ for every } A, B \subseteq S$$

The powerset 2^S of S, is a semigroup for $\cdot$ so defined.
Now, a subset P of a partial semigroup S is a *partial subsemigroup* of S, if P is a partial semigroup for $\cdot$ of S *i.e.,* $P \cdot P$ is a subset of P.
For any subset P of a partial semigroup S, following subset $P^+ = \bigcup_{n \geqslant 1} P^n$ (with $P^1 = P$ and $P^{n+1} = P^n \cdot P$ for every $n \geqslant 1$) is the smallest (for inclusion) partial subsemigroup of S containing P. Set P^+ is called the *partial semigroup generated* by P. In particular $(P^+)^+ = P^+$. Also, S is *finitely generated* if $S = P^+$ for some finite P.
A set $P \subseteq S$ is a *code* if there is no two factorization in P^+ of the same element:
$u_1 \cdots u_m = v_1 \cdots v_n \wedge u_1, \ldots, u_m, v_1, \ldots, v_n \in P \Rightarrow m = n \wedge \forall\ i \in [1 \cdots n], u_i = v_i$
A partial semigroup S is *free* if there is code P such that $P^+ = S$.
For every $W \subseteq 2^S$, we denote by $\bigcup W = \{a \mid \exists P \in W, a \in P\}$. Operator $+$ commutes with operator $\bigcup$, *i.e.,* $\bigcup(W^+) = (\bigcup W)^+$ for every $W \subseteq 2^S$.
The (left) *residual* $u^{-1}P$ of $P \subseteq S$ by $u \in S$ is following subset:
$u^{-1}P := \{v \in S \mid u \cdot v \in P\}$
and satisfies following basic equality:
$(u \cdot v)^{-1}P = v^{-1}(u^{-1}P)$ for all $u, v \in S$ and $P \subseteq S$.

Definition 2.2. Let $(S, \cdot)$ be a partial semigroup. The family $Rat(S)$ of rational subsets of S is the least family $\mathcal{R}$ of subsets of S satisfying the following conditions:

(i) $\emptyset \in \mathcal{R}; \{m\} \in \mathcal{R}$ for all m in S;
(ii) if $A, B \in \mathcal{R}$ then $A \cup B, A \cdot B$ and $A^+ \in \mathcal{R}$.

In order to generalize well known results for monoids in the case of partial semigroups, and as our purpose is to deal with graphs, we will set some notations and definitions for graphs and automata.
Let P be a subset of S. A (simple oriented labelled) *P-graph* G over V with arcs labelled in P is a subset of $V \times P \times V$. An element (s, a, t) in G is an *arc* of *source* s, *goal* t and *label* a (s and t are *vertices* of G). We denote by $Dom(G)$, $Im(G)$ and V_G the sets respectively of sources, goals and vertices of G. Each (s, a, t) of G is identified with labelled transition $s \xrightarrow[G]{a} t$ or simply $s \xrightarrow{a} t$ if G is understood.
A graph G is *deterministic* if distinct arcs with same source have distinct label: $r \xrightarrow{a} s \ \wedge \ r \xrightarrow{a} t \ \Rightarrow \ s = t$. A graph is (source) *complete* if, for every label a, every vertex is source of an arc labelled a: $\forall a \in P,\ \forall s \in V_G,\ \exists t\ s \xrightarrow{a} t$.
Set $2^{V \times P^+ \times V}$ of P^+-graphs with vertices in V is a semigroup for *composition relation*: $G \cdot H := \{r \xrightarrow{a \cdot b} t \mid \exists s, r \xrightarrow[G]{a} s \wedge s \xrightarrow[H]{b} t\}$ for any $G, H \subseteq V \times P^+ \times V$.
Relation $\xrightarrow[G^+]{u}$ denoted by $\overset{u}{\underset{G}{\Longrightarrow}}$ or simply $\overset{u}{\Longrightarrow}$ if G is understood, is the existence of a *path* in G labelled u in P^+. For any L in S, we denote by $s \overset{L}{\Longrightarrow} t$ that there exists u in L such that $s \overset{u}{\Longrightarrow} t$.
The *trace* (or set of path labels) $L(G, E, F)$ of G from a set E to a set F is the

following subset of P^+:

$$L(G, E, F) \ := \ \{u \in S \mid \exists\ s \in E,\ \exists\ t \in F,\ s \stackrel{u}{\Longrightarrow}_G t\ \}$$

Given $P \subseteq S$, a *P-automaton* A is a P-graph G whose vertices are called *states*, with an *initial state* i and a subset F of *final states*; the automaton *recognizes* subset $L(A)$ of P^+: $L(A) := L(G, \{i\}, F)$. An automaton is finite (resp. deterministic, complete) if its graph is finite (resp. deterministic, complete). This allows to state a standard result for rational subsets.

Proposition 2.3. *Given a subset P of a partial semigroup S, $Rat(P^+)$ is*

(i) the smallest subset of 2^S containing $\emptyset$ and $\{a\}$ for each $a \in P$, and closed for $\cup, \cdot, +$
(ii) the set of subsets recognized by finite P-automata,
(iii) the set of subsets recognized by finite and deterministic P-automata.

We simply translated the standards definitions of rational subsets of monoids given for example in [Be 79]. An interesting example of a partial semigroup is the subject of these lines: the set of arcs (labelled with an element of a finite set) between elements of a free monoid is a partial semigroup; its rational subsets are the rational graphs.

2.2 Partial semigroups and graphs

In this section, we will consider an important example of partial semigroup: the set of rational graphs. So consider an arbitrary finite set X and denote X^* its associated free monoid. We will consider graphs as subsets of $X^* \times \mathcal{A} \times X^*$ (the set of graphs over X^* with arcs labelled in $\mathcal{A}$). For convenience, set $2^{X^* \times \mathcal{A} \times X^*}$ is denoted $G_{\mathcal{A}}(X^*)$.
Now, with $(u, a_i, v) \cdot_i (u', a_i, v') = (u \cdot u', a_i, v \cdot v')$, set $X^* \times \{a_i\} \times X^*$ (a_i in $\mathcal{A}$) is a monoid. As stated in Example 2.1 the union of these monoids (namely $X^* \times \mathcal{A} \times X^*$) is a partial semigroup. We denote by $\cdot$ the operation in $X^* \times \mathcal{A} \times X^*$ (which is $\cdot_i$ for each $X^* \times \{a_i\} \times X^*$).
Remark: this $\cdot$ operation for graphs is indeed, similar to the synchronization product for transition systems defined by Nivat and Arnold in [AN 88].
We are now able to define the set of rational graphs.

Definition 2.4. The set of rational graphs, denoted $Rat(X^* \times \mathcal{A} \times X^*)$ is the family of rational subsets of $X^* \times \mathcal{A} \times X^*$.

Let us now recall that a *transducer* is a finite automaton over pairs (see for example [Au 88] [Be 79]). A rational relation (*i.e.,* a rational subset of $X^* \times X^*$) is recognized by a rational transducer.
There is a strong relationship between rational graphs and rational relations and to characterize the family of rational graphs in a more practical way we will use *labelled transducers*.

Definition 2.5. A *labelled transducer* $T = \langle Q, I, F, E, L\rangle$ over X, is composed of a finite set of states Q, a set of initial states $I \subseteq Q$, a set of final states $F \subseteq Q$, a finite set of transitions (or edges) $E \subseteq Q \times X^* \times X^* \times Q$ and an application L from F into $2^{\mathcal{A}}$.

Like for P-graphs, transition (p, u, v, q) of transducer T will be denoted by $p \xrightarrow[T]{u/v} q$ or simply $p \xrightarrow{u/v} q$ if T is understood. Now similarly an element $(u, d, v) \in X^* \times \mathcal{A} \times X^*$ is *recognized* by transducer T if there is a path $p_0 \xrightarrow[T]{u_1/v_1} p_1 \cdots p_{n-1} \xrightarrow[T]{u_n/v_n} p_n$ and $p_0 \in I$, $p_n \in F$, $u = u_1 \cdots u_n$, $v = v_1 \cdots v_n$ and $d \in L(p_n)$.
Remark: an illustration of transducer execution will be given in Example 2.7.

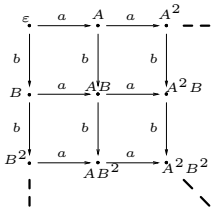
Proposition 2.6. *A graph G in $G_{\mathcal{A}}(X^*)$ is rational if and only if it satisfies one of the following equivalent properties:*

(i) G belongs to the smallest subset of $G_{\mathcal{A}}(X^*)$ containing:
$\emptyset$,$\{\varepsilon \xrightarrow{d} \varepsilon\}$,$\{x \xrightarrow{d} \varepsilon\}$ and $\{\varepsilon \xrightarrow{d} x\}$, for all $x \in X$, all $d \in \mathcal{A}$, and closed under $\cup$, $\cdot$ and $+$;
(ii) G is a finite union of rational relations over each letter:
$G = \bigcup_{d\in\mathcal{A}} R_d$, for $R_d \in Rat(X^* \times \{d\} \times X^*)$;
(iii) G is recognized by labelled rational transducer.

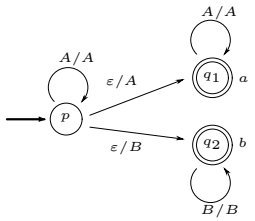
This Proposition states that for any graph G in $Rat(X^*\times\mathcal{A}\times X^*)$, the relation: $\xrightarrow[G]{d} := \{(u, v) \mid u \xrightarrow[G]{d} v\}$ is rational for each d in $\mathcal{A}$. Therefore we also introduce $\xrightarrow[G]{} := \bigcup_{d\in\mathcal{A}} \xrightarrow[G]{d}$, which is also a rational relation. Naturally we denote by $\xrightarrow[G]{d}(u)$ (resp. $\xrightarrow[G]{}(u)$) the image of word u by relation $\xrightarrow[G]{d}$ (resp. $\xrightarrow[G]{}$) (and similarly for subsets of X). Also for a rational graph G there are possibly many transducers generating it, thus we will denote by $\Theta(G)$ the set of transducers generating G.

We will now give some examples of rational graphs.

Example 2.7. This graph :



is a rational graph generated by this transducer :

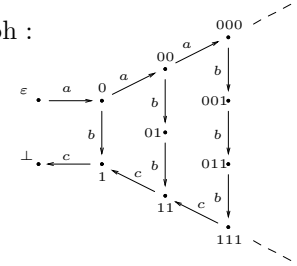


Notice that its second order monadic theory is undecidable and therefore rational graphs have an undecidable second order monadic theory.
Why does the arc (AB, b, AB^2) belong to the graph? Simply because the following path is in the transducer:

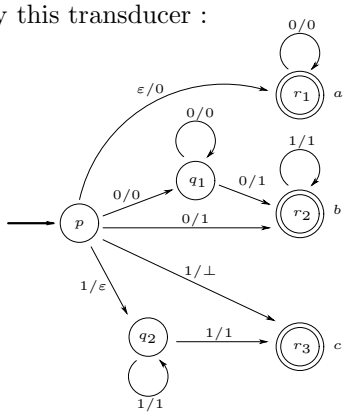
$$p \xrightarrow{A/A} p \xrightarrow{\varepsilon/B} q_2 \xrightarrow{B/B} q_2$$

and that b is associated to the final state q_2.

Example 2.8. This graph :

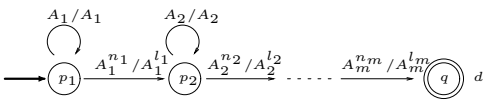


is rational, generated by this transducer :



We finish with a last example showing that the transition graphs of *Petri nets* are rational graphs.

Example 2.9. For more detail on Petri nets the reader may refer to [Re 85]. A Petri net can be seen as a finite set of transitions of this form:
$A_1^{n_1} A_2^{n_2} \cdots A_m^{n_m} \xrightarrow{d} A_1^{l_1} A_2^{l_2} \cdots A_m^{l_m}$, with A_i^x representing there are x coins in place A_i (d represents the label (if any) of the transition). Following transducer generates the transition graph associated to the above transition:



Each vertex of the generated graph correspond to a marking of the Petri net. Each arc of the graph represents that a transition has been fired.

2.3 Some results for rational graphs

This section will introduce results for this family of graphs. Some of these results are just a reformulation of known results over rational relations. Others are simple facts on these graphs and their boundary.
The first fact is that this family is an extension of previous families. Simply recall that every *prefix-recognizable graph* (defined in [Ca 96]) is a finite union of graphs of the following form :

$$(U \xrightarrow{a} V) \cdot W := \{ uw \xrightarrow{a} vw \mid u \in U \wedge v \in V \wedge w \in W \}$$

with U, V, W rational sets.
This characterization ensures that *prefix-recognizable graphs* are rational graphs. As the regular graphs (defined in [Co 90]) are *prefix-recognizable graphs*, they are rational too. Furthermore, the graphs in Examples 2.7 and 2.8 are not *prefix-recognizable graphs* thus the inclusion is strict. Let us now translate some well-known results for rational relations, to rational graphs (the proofs will be omitted they are mostly direct consequences of results found in [Au 88] and [Be 79]).

Proposition 2.10. *A rational graph G is of finite out-degree if and only if there exists a transducer $T \in \Theta(G)$ such that there exists no cycle in T labelled on the left with the empty word which is not labelled on the right with the empty word. In other words the only cycles labelled on the left ε, are labelled on the right ε.*

Remark: naturally this proposition can be translated to characterize the graphs of finite in-degree, by simply replacing right by left and vice-versa.

Proposition 2.11. *Every rational graph is recursive: it is decidable whether an arc (u, d, v) belongs to a rational graph.*

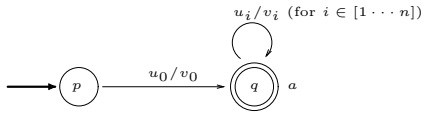
Theorem 2.12. *It is decidable whether a rational graph is deterministic (from its transducer).*

Proposition 2.13. *The inclusion and equality of deterministic rational graphs is decidable.*

Remark: unfortunately this result ceases to be true for general rational graphs ([Be 79] Theorem 8.4, page 90).
We have already seen that the second order monadic theory of these graphs is undecidable in general. We will now see that it is also the case for the first order theory.

Proposition 2.14. *The first order theory of rational graphs is undecidable.*

Proof. We will prove this proposition by reducing Post's correspondence problem (P.C.P.) to this problem. Let us recall the P.C.P.: given an alphabet X and (u_0, v_0), (u_1, v_1),..., (u_n, v_n) elements of $X^* \times X^*$. Does there exist a sequence $0 \leqslant i_1, i_2, \ldots, i_m \leqslant n$, such that $u_0 u_{i_1} \cdots u_{i_m} = v_0 v_{i_1} \cdots v_{i_m}$? To an instance of P.C.P. (i.e. a family (u_i, v_i)) we associate following transducer:



The resolution of P.C.P. becomes finding a vertex s such that $s \xrightarrow{a} s$ is an arc of the graph generated by the transducer. It is a first order instance, therefore,as P.C.P. is undecidable, the first order theory of rational graphs is not decidable in general. □

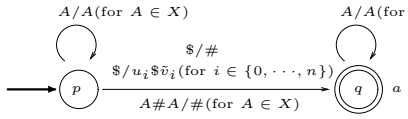
Before giving another negative decision result, let us denote by $\tilde{u}$ the mirror of word u (defined by induction on the length of u: $\tilde{\varepsilon} = \varepsilon$ and $\widetilde{au} = \tilde{u}a$ (for any u with $|u| \geqslant 0$).

Proposition 2.15. *Accessibility is not decidable for rational graphs in general.*

Proof. Once again, we use P.C.P. Using the same notations as earlier define a (word) rewriting system G, using two new symbols $\#$ and \$, in the following way:

$$G \begin{cases} \$ & \longrightarrow u_i \$ \widetilde{v_i} \ \forall i \in \{0, \cdots, n\} \\ \$ & \longrightarrow \ \# \\ A\#A & \longrightarrow \ \# \quad \forall A \in X \end{cases}$$

Now "P.C.P. has a solution" is equivalent to the existence of a derivation from $u_0\$\widetilde{v_0}$ to $\#$. But, considering the following transducer:



the question becomes: is there a path leading from $u_0\$\widetilde{v_0}$ to the vertex $\#$? Answering the last question would allow P.C.P. to be solved in the general case which is a contradiction. Therefore accessibility is undecidable for the rational graphs in general. □

Remark: the transitive closure of a rational graph is, at least, uneffective. If this construction were effective and rational, then accessibility for rational graph would be decidable.
Now we will see a case where accessibility is decidable for rational graphs. A transducer T is *increasing* if every pair (u, v) recognized by T is such that the length of v (denoted by $|v|$) is greater or equal to the length of u : $|v| \geqslant |u|$.

Proposition 2.16. *The accessibility is decidable for any rational graph with an increasing transducer.*

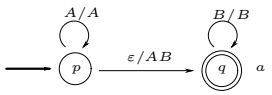
Proof. Let us denote by $T^{\leqslant n}(u)$ following set: $T^{\leqslant n}(u) := \bigcup_{i=0}^{n} T^i(u)$. For all $n \in \mathbb{N}$ this set is rational.
Now, let G be a rational graph generated by an increasing transducer T and let u and v be two vertices of G. Let us put $n_0 = |\{w \in X^* \mid |u| \leqslant |w| \leqslant |v|\}| = |X|^{|u|} + \cdots + |X|^{|v|}$. Vertex v is accessible from u if and only if v belongs to $T^{\leqslant n_0}(u)$. Thus accessibility is decidable for rational graphs with an increasing transducer. □

We now give a technical Lemma that allows the construction of a graph that is not *structurally* rational.

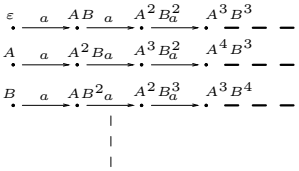
Lemma 2.17. *Let G be a rational graph of finite out-degree. There exists two integers p and q such that for every $(s, a, t) \in G$ we have $|t| \leqslant p.|s| + q$*

Example 2.18. Consider an infinite tree in $X^* \times \mathcal{A} \times X^*$ such that every vertex of depth n has $2^{2^{2^n}}$ sons. This tree is not *strucurally* rational, in other words whatever name are given to its vertices this graph is never a rational graph. This is a direct consequence of previous lemma: say n is the length of the root, there are at most $|X|^{(np^l + p^{l-1}q + \cdots + q)}$ vertices of depth l.

Despite these results the transducers are not able to capture the structure of rational graphs.
For example, this transducer:



generates this graph:



The connected component of the empty word, ε, is a straight-line. It is "up to isomorphism" obviously rational, but as a sub-graph of this graph, it is not rational (its vertices form a context-free language). Therefore we need an external ("up to isomorphism") characterization of these graphs. This is the subject of the next section.

3 An external characterization

In this section, we will characterize rational graphs using inverse linear substitutions. Labelled transducers are an internal representation of rational graphs, it clearly depends on the name of the vertices. But often in graph theory, the name of the vertices is not relevant, it carries no information. An external characterization, like the graph grammars for equational graphs, produces graphs without giving names for vertices. It only gives the *structure* of the graph. Inverse linear substitution is an external characterization of rational graphs.

3.1 Graph isomorphism

An external characterization of rational graphs is given "up to isomorphism".
Two graphs G_1 and G_2 in $G_{\mathcal{A}}(X^*)$ are *isomorphic*, if there is a bijection $\psi : V(G_1) \to V(G_2)$ such that: $s_1 \xrightarrow[G_1]{d} s_2$(*i.e.,* $(s_1, d, s_2) \in G_1$) if and only if $\psi(s_1) \xrightarrow[G_2]{d} \psi(s_2)$.

Two isomorphic graphs have the same structure: they are the same up to a renaming of the vertices.
Now let us consider the equivalence ($\equiv$) generated by graph isomorphism: we say that G_1 is equivalent to G_2 (denoted $G_1 \equiv G_2$) if G_1 and G_2 are isomorphic. This equivalence relation provides us with a partition of $G_{\mathcal{A}}(X^*)$ denoted $Graph_{\mathcal{A}} := G_{\mathcal{A}}(X^*)/\equiv$. This allows the introduction of the set of *structural rational graphs*:

$$GRat_{\mathcal{A}} := \{[G]_{\equiv} \in Graph_{\mathcal{A}} \mid G \in Rat(X^* \times \mathcal{A} \times X^*)\}$$

This set is the set of graphs that are isomorphic to some rational graph.
Set $Graph_{\mathcal{A}}$ (and $GRat_{\mathcal{A}}$) does not depend on the choice of set X, therefore we can choose X to be any two letters alphabet with no loss of generality.

Lemma 3.1. *For all subset X' (with at least two elements) of X and all class $[G]_{\equiv}$ of $Graph_{\mathcal{A}}$ $(= G_{\mathcal{A}}(X^*)/\equiv)$ there exists G_0 in $G_{\mathcal{A}}(X'^*)$ such that $G_0 \in [G]_{\equiv}$.*

We now have to characterize the structure of $GRat_{\mathcal{A}}$. This is the goal of the next section.

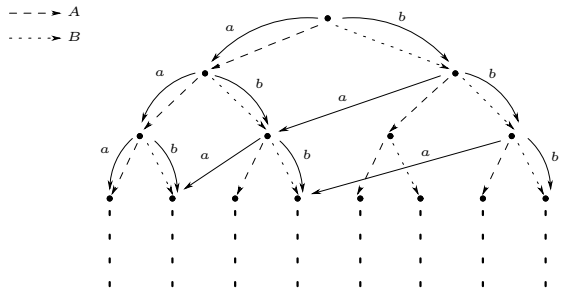
3.2 Substitution

Recall the definition of the *prefix-recognizable graphs* (family REC_{Rat}). This family has been defined as the set of graphs obtained from the complete binary tree by inverse rational substitution, followed by rational restriction. We will use the same process (actually a linear context-free substitution) to obtain the family of rational graphs.
A *substitution* over a free monoid X^* is a morphism $\varphi : \mathcal{A}^* \to 2^{X^*}$, which associates to each letter in $\mathcal{A}$ a language in X^*. Our purpose is to study graphs, starting from the complete binary tree (Λ) labelled $X = \{A, B\}$. To move by inverse arcs, we use a new alphabet : $\overline{X} = \{\overline{A}, \overline{B}\}$ and we say that $x \xrightarrow{\overline{A}} y$ if $y \xrightarrow{A} x$. Given a language L and two vertices x and y, recall that $x \underset{\Lambda}{\overset{L}{\Longrightarrow}} y \Leftrightarrow \exists u \in L, x \underset{\Lambda}{\overset{u}{\Longrightarrow}} y$. Now, given a substitution $\varphi : \mathcal{A}^* \to 2^{(X \cup \overline{X})^*}$, we can define the graph $\varphi^{-1}(\Lambda)$ in the following way:

$$\varphi^{-1}(\Lambda) = \{x \xrightarrow{d} y \mid d \in \mathcal{A} \wedge \ x \underset{\Lambda}{\overset{\varphi(d)}{\Longrightarrow}} y\}$$

Given a language L, we define now $L_\Lambda = \{s \mid r \underset{\Lambda}{\overset{L}{\Longrightarrow}} s\}$. It allows us to consider the graph $\varphi^{-1}(\Lambda)_{|L_\Lambda}$: it is the image of the complete binary tree by an inverse substitution followed by a restriction; if L is rational, we say a rational restriction.

Example 3.2. Example 2.7 states that the grid is a rational graph. Following substitution: $h(a) = \{\overline{B}^m A B^m \mid m \geqslant 0\}$, $h(b) = \{B\}$ over the complete binary tree on $\{A, B\}$, followed with the restriction to $L = A^*B^*$ produces a graph isomorphic to the grid:



Now, it is well know that there is a close relationship between *linear languages* and rational relations (a linear language is a context-free language generated by a grammar with only, at most, one non-terminal on the right hand side of each rule). And indeed, if we denote the set of linear languages over the alphabet $X \cup \overline{X}$ by $Lin(X \cup \overline{X})$, we have the following proposition.

Proposition 3.3. *The set $GRat_{\mathcal{A}}$ is a subset of the family of the graphs obtained from the complete binary tree (Λ) by an inverse* linear *substitution, followed by a rational restriction:*

$$GRat_{\mathcal{A}} \subseteq \{[\varphi^{-1}(\Lambda)_{|L_\Lambda}]_{\equiv} \mid \forall d \in \mathcal{A}, \varphi(d) \in Lin(X \cup \overline{X}) \wedge L \in Rat(X)\}$$

Proof (Sketch). We first transform the transducer generating the graph (G) so that each vertex begins with the same prefix. Then we produce linear languages (L_d) such that $(u, d, v) \in X^* \times \mathcal{A} \times X^*$ is an arc of G if and only if $\widetilde{\overline{u}}v \in L_d$. We then define $\varphi(d)$ to be L_d. It only remains to define L (the rational restriction) to be $L := Dom(G) \cup Im(G)$ □

The converse of this result would help us to grab the structure of rational graphs. Unfortunately it is not obvious. Actually the following example illustrate the difficulty of the naive converse of Proposition 3.3.

Example 3.4. Consider $\varphi(a) = \{\overline{B}BA^nB^n | \, n \in \mathbb{N}\}$, it is a linear substitution. Consider $L = BA^*B^*$ and the graph $G = \varphi^{-1}(\Lambda)_{|L_\Lambda}$. Structurally, graph G is rational (it is the star). But the graph naturally associated to G (according to $\varphi(a)$ and L) is $G' = \{(B, a, BA^nB^n) | \, n \in \mathbb{N}\}$, which is not rational.

So there is a deep isomorphism problem to get the converse. Actually, we will try to inject rationality in the "linear language" to achieve a complete characterization of rational graphs.

A natural way to introduce rationality into $Lin(X \cup \overline{X})$ would be to impose the projections over barred and non-barred letters to be rational. The next example shows that again, things are not so nice.

Example 3.5. Consider $\varphi(a) = \{\overline{A}\,\overline{B}BA^nB^m | \; n \geqslant m\} \cup \{\overline{B}BA^nB^m | \; m > n\}$ φ is a linear substitution. Moreover it has rational projections over barred and non-barred letters. Consider $L = BA^*B^*$ and the graph $G = \varphi^{-1}(\Lambda)_{|L_\Lambda}$.

Structurally, graph G is rational (it is two stars). But the graph naturally associated to G (according to $\varphi(a)$ and L) is $G' = \{(BA, a, BA^nB^m) |\ n \geqslant m\} \cup \{(B, a, BA^nB^m) |\ m > n\}$, which is not rational (its intersection with the recognizable set $\{BA\} \times \{a\} \times BA^*B^*$ is $\{(BA, a, BA^nB^m) |\ n \geqslant m\}$ which is not rational).

Now consider the set $Ratlin(X \cup \overline{X})$ of linear languages (called rational-linear) over $(X \cup \overline{X})^*$ such that the production of their grammars are of following form: $p \to \overline{u}qv$ (with $\overline{u} \in \overline{X}^*$ and $v \in X^*$) or $p \to \varepsilon$.

Theorem 3.6. *Set $GRat_{\mathcal{A}}$ is precisely the set of graphs obtained from the complete binary tree (Λ) by a* rational-linear *substitution, followed by a rational restriction :*

$$GRat_{\mathcal{A}} = \{[\varphi^{-1}(\Lambda)_{|L_\Lambda}]_{\equiv} \mid \forall d \in \mathcal{A}, \varphi(d) \in Ratlin(X \cup \overline{X}) \wedge L \in Rat(X)\}$$

Proof (Sketch). The first inclusion is treated in Proposition 3.3. For the reverse inclusion we first take a graph G image of a *rational-linear* substitution, followed by a rational restriction then we need to check that it is possible to produce a transducer from the grammars of $\varphi(d)$ for each d. Then we show that this graph contains G, finally, using rational intersection we obtain precisely G. □

Now that an external characterization of the rational graphs has been given, the next section will consider the properties of the *traces* of rational graphs.

4 The traces of rational graphs

We have already seen that there is a strong connection between language theory and rational graphs. In this section we will see another connection between graphs and languages, in terms of *traces*.
We first recall that the *trace* of a graph G leading from a vertex set I (of initial states) to a vertex set F (of final states) is the set of all the path labels in the graph, leading from a vertex in the set of initial states to a vertex in the set of final states:

$$L(G, I, F) := \{u \mid \exists s \in I\ \exists t \in F, s \xRightarrow[G]{u} t\}$$

In other words the trace of a graph is "the language of its labels". For example the traces of the finite graphs are all rational languages and the traces of *prefix-recognizable graphs* are all context-free languages. Notice by the way that the traces of rational graphs contain therefore every context free language.

Proposition 4.1. *The traces of rational graph leading from a rational vertex set to a context free vertex set (or vice-versa) is recursive*

Proof. In order to check whether a word u is in the trace of graph G (from a set I to a set F), it is just to check if the set $S = \xrightarrow[G]{u(|u|)}(\xrightarrow[G]{u(|u|-1)}(\cdots \xrightarrow[G]{u(1)}(I)\cdots))$

intersects set F. If set I is rational (*resp.* context-free) its image by a rational transduction is rational (*resp.* context-free), hence by a simple induction, set S is rational (*resp.* context-free). Therefore it is decidable whether $S \cap F$ is empty. □

Let us denote by TR the family of the traces of rational graphs leading from a rational vertex set to a rational vertex set: $TR = \{L(G, I, F) | \, G \in Rat(X^* \times \mathcal{A} \times X^*) \wedge I, F \in Rat(X^*)\}$ (notice that we could as well restrict ourselves to a unique initial state and a unique final state). Now we will show that set TR form an *Abstract Family of Languages* (AFL), that is, it satisfies following properties:

- closure for intersection with a rational (regular) language,
- closure under non-erasing (monoid)morphism, and inverse morphism,
- for each $L, L' \in TR$ we have $L \cdot L', L \cap L', L^+, L^* \in TR$.

Proposition 4.2. *The intersection of two elements of TR is an element of TR.*

Proof. Consider two elements L and L' of TR. Say $L = L(G, I, F)$ and $L' = L(G', I, F)$. The language $L \cap L'$ is actually the trace of $G \cdot (\{\$\} \times \mathcal{A} \times \{\$\}) \cdot G'$ (with $\$$ a new symbol) between $I_G \cdot \{\$\} \cdot I_{G'}$ and $F_G \cdot \{\$\} \cdot F_{G'}$. Hence $L \cap L'$ in an element of TR. □

As rational languages are traces of rational graphs (finite graphs are rational graphs), family TR is closed under intersection with rational languages.
Now let us recall that a finite (*resp.* rational) substitution $\sigma : \mathcal{A}^* \to 2^{\mathcal{A}^*}$ is a morphism such that for each letter d in $\mathcal{A}$ $\sigma(d)$ is a finite (*resp.* rational) subset of $\mathcal{A}^*$. A substitution is *non-erasing* if $\varepsilon \notin \sigma(d)$ for all $d \in \mathcal{A}$.

Proposition 4.3. *Family TR is closed under non-erasing finite substitution.*

Proof (Sketch). Consider σ a non-erasing finite substitution, and L a language in TR. We take a graph G such that $L = L(G, I, F)$, and T a transducer generating G. We, then, construct a new transducer such that each production d in T is replaced by a path u (in the corresponding graph), for each $u \in \sigma(d)$. The trace of the graph generated by this transducer is $\sigma(L)$ □

Following corollary is a direct consequence of this proposition.

Corollary 4.4. *Family TR is closed under non-erasing morphism.*

Notice that the condition "non-erasing" is essential for our proof. A interesting question is whether this condition is necessary.

Proposition 4.5. *Assume that L is an element of TR and that σ is a finite substitution over $\mathcal{A}^*$ then $\sigma^{-1}(L)$ is a language of TR.*

Proof (Sketch). This proposition is a consequence of Elgot and Mezei's theorem, which states that the composition of two rational relations is a rational relation (see for example [Be 79], Theorem 4.4 p 68). Using this we can produce a rational graph in which a finite number of finite path are replaced by arcs. Which proves the Proposition. □

Remark: Note that it is not as straightforward for inverse rational substitution. Actually it seems that it is not true for inverse rational substitution: consider any rational graph with one label (a) and the inverse rational substitution $\sigma(a) = a^*$. The graph image with the same approach would be the transitive closure of the original graph, which is not effectively rational (and might not even be structurally rational) as stated in the remark after Proposition 2.15.
Following corollary is an obvious consequence of proposition 4.5.

Corollary 4.6. *Family TR is closed under inverse morphism.*

Proposition 4.7. *Family TR is closed under concatenation, Kleene plus and star.*

Proof (Sketch). The argument is more or less the same as for finite automata. We use operation over rational relations to get the results. □

As stated earlier, we only have now to summary these results.

Theorem 4.8. *The traces of rational graphs, leading from a rational vertex set to a rational vertex set, form an AFL (Abstract Family of Languages).*

Proof. This result is simply a brief summary of corollaries 4.4, 4.6 and propositions 4.2 and 4.7. □

Now we have an abstract family of languages that contains the context free languages. This AFL is a subset of the recursive languages. It seems that this family is composed of the context sensitive languages.

Conjecture 4.9. *The traces of the rational graphs are precisely the context sensitives languages.*

Notice also that recently graphs of linear bounded machines (which characterize context sensitive languages) have been studied in [KP 99].

5 Conclusion

In this paper, a general family of graphs has been introduced. Rational graphs are a strict extension of previously studied families. It is a well grounded family, related to well known structures of language theory. We have given both an internal and an external characterization, as well as some basic properties.

Unfortunately, or fortunately depending on the point of view, it is a very expressive family. Therefore many decision results are lost. An interesting question is to study restrictions of this family that will retain decision results from former families.

Traces of rational graphs are another aspect of this family. We have shown that it forms an abstract family of recursive languages. An interesting question is to know if these traces are precisely the context sensitive languages.

Rational trees also seem to be an interesting field of research, but this has not been done yet.

Acknowledgements

The author would like to express his gratitude to Didier Caucal for his help along the preparation of this paper.

References

[AN 88] A. Arnold and M. Nivat, *Comportements de processus*, Colloque AFCET "les mathématiques de l'informatique", pp. 35–68, 1982.

[Au 88] J.-M. Autebert and L. Boasson, *Transductions rationelles*, Ed. Masson, pp. 1–133, 1988.

[Be 79] J. Berstel *Transductions and context-free languages*, Ed. Teubner, pp. 1–278, 1979.

[Ca 96] D. Caucal *On transition graphs having a decidable monadic theory*, LNCS 1099, pp. 194–205, 1996,

[Co 90] B. Courcelle *Graph rewriting: an algebraic and logic approach*, Handbook of TCS, Vol. B, Elsevier, pp. 193–242, 1990.

[GG 66] S. Ginsburg and S. A. Greibach *Mappings which preserve context sensitive languages*, Information and Control 9, pp. 563–582, 1966.

[HU 79] J. E. Hopcroft and J. D. Ullman *Introduction to automata theory, langages and computation*, Ed. Addison-Wesley pp. 1–284, 1979.

[KP 99] T. Knapik and E. Payet *Synchronization product of Linear Bounded Machines*, LNCS 1684, pp. 362–373, 1999.

[MS 85] D. Muller and P. Schupp *The theory of ends, pushdown automata, and second-order logic*, TCS 37, pp. 51–75, 1985.

[Ni 68] M. Nivat *Transduction des langages de Chomsky*, Ann. de l'Inst. Fourier 18, pp. 339-456, 1968.

[Re 85] W. Reisig *Petri nets*. EATCS Monographs on Theoretical Computer Science, Vol. 4, Springer Verlag, 1985.

[Sc 76] M.P. Schützenberger *Sur les relations rationnelles entre monoïdes libres*, TCS 3, pp. 243–259, 1976.