Deterministically and Sudoku-deterministically recognizable 2-dimensional languages

Bernd Borchert and Klaus Reinhardt

University of Tübingen

Contents

- Characterizations of regular languages
- Logic characterization of picture languages
- Definition of recognizable picture languages REC
- Deterministic versions: $SDREC$, $DREC$
- $SDREC \supseteq 4AFA$ contains languages outside REC
- $DREC \subset REC$
- Directed acyclic graphs in REC by elimination of souces and sinks

Characterizations of regular languages:

Regular grammars,

regular expressions,

(non-)deterministic finite automata,

finite monoids,

tiling systems,

and **(existential) monadic second-order logic**.

Example: The language $(ab)^*a$ is described by the formula

$$\forall x\ (\ ((\neg\exists y = x-1 \vee \neg\exists y = x+1) \longrightarrow Q_a(x)) \wedge$$
$$(\forall y = x+1 \longrightarrow ((Q_a(x) \wedge Q_b(y)) \vee (Q_b(x) \wedge Q_a(y))))\ ).$$

$$\_ababababababa\_$$

Example: The language $(ab)^*a$ is described by the formula

$$\forall x\ (\ ((\neg\exists y = x-1 \vee \neg\exists y = x+1) \rightarrow Q_a(x)) \wedge$$
$$(\forall y = x+1 \rightarrow ((Q_a(x) \wedge Q_b(y)) \vee (Q_b(x) \wedge Q_a(y))))\ ).$$

$ababababababababa$

Example: The language $(ab)^*a$ is described by the formula

$$\forall x\ (\ ((\neg\exists y = x-1 \vee \neg\exists y = x+1) \rightarrow Q_a(x)) \wedge$$

$$(\forall y = x+1 \rightarrow ((Q_a(x) \wedge Q_b(y)) \vee (Q_b(x) \wedge Q_a(y))))\ ).$$

$$abababababababa$$

Example: The language $(ab)^*a$ is described by the formula

$$\forall x \; (\; ((\neg\exists y = x-1 \vee \neg\exists y = x+1) \rightarrow Q_a(x)) \wedge$$

$$(\forall y = x+1 \rightarrow ((Q_a(x) \wedge Q_b(y)) \vee (Q_b(x) \wedge Q_a(y)))) \;).$$

The language $(aa)^*a$ is described by the second-order formula

$$\exists X \; \forall x \; (\; ((\neg\exists y = x-1 \vee \neg\exists y = x+1) \rightarrow X(x)) \wedge \forall x \; Q_a(x)) \wedge$$

$$\forall y = x+1 \; (X(x) \leftrightarrow \neg X(y)) \;).$$

$$aaaaaaaaaaa = \pi(ababababababa) \text{ with } \pi(a) = \pi(b) = a$$

Remark: The size of a finite automaton can be non elementary in the size of the corresponding formula [Rei02].

Pictures

A *picture* over Σ is a two-dimensional array of elements of Σ.

A *picture language* is a set of pictures $\subseteq \Sigma^{**}$.

Recognizable picture languages [GR92]

[GRST94]: Characterization by **existential monadic second order logic** using horizontal and vertical neighbor relations H and V.

Example: The language of pictures p over $\{a\}$ having size $(2^k, k)$ is recognizable

a	a	a	a	a	a	a	a
a	a	a	a	a	a	a	a
a	a	a	a	a	a	a	a

by a projection π with $\pi(0) = \pi(1) = a$ from the language of pictures p over $\{0,1\}$ having size $(2^k, k)$

0	1	0	1	0	1	0	1
0	0	1	1	0	0	1	1
0	0	0	0	1	1	1	1

such that the i-th column of p is the binary representation of $i-1$.

0	1	0	1	0	1	0	1
0	0	1	1	0	0	1	1
0	0	0	0	1	1	1	1

The language of pictures p over $\{0,1\}$ having size $(2^k, k)$ such that the i-th column of p is the binary representation of $i-1$ is described by the first-order formula

$$\forall x\ (((\neg\exists y\ H(y,x)) \longrightarrow Q_0(x)) \wedge (((\neg\exists y\ H(x,y)) \longrightarrow Q_1(x)) \wedge$$

$$\forall x,y\ (H(x,y) \longrightarrow (\ ((\exists z,v\ (V(z,x) \wedge V(v,y) \wedge Q_1(z) \wedge Q_0(v))) \vee (\neg\exists z\ V(z,x))) \longrightarrow$$

$$((Q_0(x) \wedge Q_1(y)) \vee (Q_1(x) \wedge Q_0(y)))\ ) \wedge$$

$$((Q_0(x) \vee Q_1(y)) \longrightarrow \exists z,v$$

$$(V(x,z) \wedge V(y,v) \wedge ((Q_0(z) \wedge Q_0(v)) \vee (Q_1(z) \wedge Q_1(v))))\ )\ ))$$

Recognizable picture languages [GR92] [GRST94]

A *picture* over Σ is a two-dimensional array of elements of Σ.

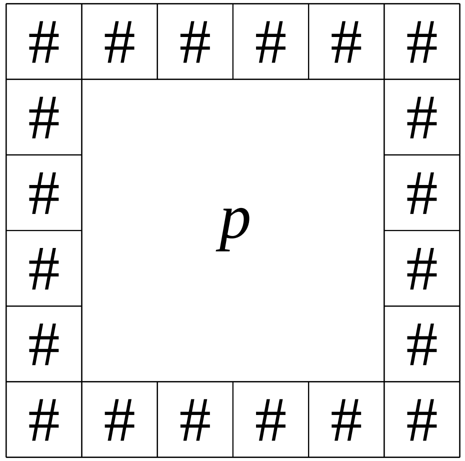
A *picture language* is a set of pictures $\subseteq \Sigma^{**}$.

For a $p \in \Sigma^{**}$ of size (m, n),

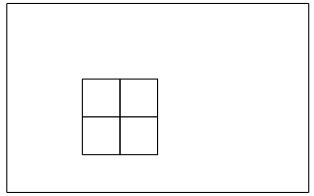
$\hat{p}$ has size $(m+2, n+2)$ adding

a frame of symbols $\# \notin \Sigma$.

$\hat{p} :=$



Let $T_{2,2}(p)$ be the set of all subpictures of p with size $(2,2)$.



Local picture language: $\mathscr{L}(\Delta) := \{p \in \Gamma^{*,*} | T_{2,2}(\hat{p}) \subset \Delta\}$.

Example: The language of pictures p over $\{0,1\}$ having size $(2^k, k)$ such that the i-th column of p is the binary representation of $i-1$.

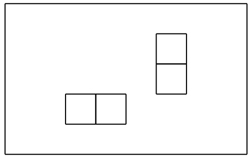
$\Delta = T_{2,2}(\hat{p})$ with $\hat{p} :=$

#	#	#	#	#	#	#	#	#	#
#	0	1	0	1	0	1	0	1	#
#	0	0	1	1	0	0	1	1	#
#	0	0	0	0	1	1	1	1	#
#	#	#	#	#	#	#	#	#	#

Recognizable picture language: $\mathscr{L}(\Delta,\pi) := \pi(\mathscr{L}(\Delta)) \in REC$

Example: By π with $\pi(0) = \pi(1) = a$, the language of pictures p over $\{a\}$ having size $(2^k, k)$ is recognizable

#	#	#	#	#	#	#	#	#	#
#	a	a	a	a	a	a	a	a	#
#	a	a	a	a	a	a	a	a	#
#	a	a	a	a	a	a	a	a	#
#	#	#	#	#	#	#	#	#	#

Let $\Sigma \cap \Gamma = \emptyset$, $\pi : \Gamma \to \Sigma$ and $\Delta \subseteq (\Gamma \cup \{\#\})^{1,2} \cup (\Gamma \cup \{\#\})^{2,1}$, which means we consider two kinds of tiles: 

$\mathscr{L}_{hv}(\Delta) := \{p \in \Gamma^{*,*} | T_{1,2}(\hat{p}) \cup T_{2,1}(\hat{p}) \subseteq \Delta\}$ is called *hv-local*

[LS97]: Every language in $\in REC$ can be written as $\pi(\mathscr{L}_{hv}(\Delta))$

Theorem [Rei00] The language of pictures, where the number of a's is equal to the number of b's and having a size (n,m) with $\log n \leq m \leq 2^n$ is recognizable.

Theorem [Rei98] The language of pictures over $\{a,b\}$, where all occurring b's are connected is recognizable.

Idea: Guess and locally check a tree of b's. (w.l.o.g. being rooted at the lowest b on the left side).

Difficulty: Avoid cycles.

Recognition of a picture as a deterministic process

[AGMR06]: $CR{-}DREC \subset UREC \subset REC$

[Rei98]: Rules to derive the pre-immage symbols locally

Given a picture p over Σ we initialize every position (i,j) by the set $s_p(i,j) := \pi^{-1}(p(i,j)) \in 2^{\Gamma}$ of possible pre-image symbols.

On $(2^\Gamma)^{*,*}$ we allow steps $s \underset{sd(\Delta,\pi)}{\Longrightarrow} s'$ where for all i,j we have $s'(i,j) =$

$$\{x \in s(i,j) \mid \exists y \in s(i+1,j), \boxed{x \mid y} \in \Delta \wedge \exists y \in s(i-1,j), \boxed{y \mid x} \in \Delta \wedge$$
$$\exists y \in s(i,j+1), \begin{array}{|c|}\hline x \\ \hline y \\ \hline\end{array} \in \Delta \wedge \exists y \in s(i,j-1), \begin{array}{|c|}\hline y \\ \hline x \\ \hline\end{array} \in \Delta\}.$$

The definition in [Rei98] can be formulated in similar terms:

On $(2^\Gamma)^{*,*}$ we allow steps $s \underset{d(\Delta,\pi)}{\Longrightarrow} s'$ where for all i,j we have $s'(i,j) =$

$$\{x \in s(i,j) \mid \exists y \in s(i+1,j), \boxed{x \mid y} \in \Delta \wedge \exists y \in s(i-1,j), \boxed{y \mid x} \in \Delta \wedge$$
$$\exists y \in s(i,j+1), \begin{array}{|c|}\hline x \\ \hline y \\ \hline\end{array} \in \Delta \wedge \exists y \in s(i,j-1), \begin{array}{|c|}\hline y \\ \hline x \\ \hline\end{array} \in \Delta\}$$

if $|s'(i,j)| = 1$ and otherwise $s'(i,j) = s(i,j)$.

Accepted language: $\mathscr{L}_{(s)d}(\Delta,\pi) :=$ $\{p \in \Sigma^{*,*} | \hat{s}_p \underset{(s)d(\Delta,\pi)}{\overset{*}{\Longrightarrow}} \hat{s}'$ with $s'(i,j) = \{p'(i,j)\}$ for all i,j and $p' \in \mathscr{L}(\Delta)\}$.

The class $(S)DREC$ is the set of picture languages $L \subseteq \Sigma^{*,*}$ which are *(Sudoku-)deterministically recognizable*, that means there are Δ, π with $L = \mathscr{L}_{(s)d}(\Delta,\pi)$.

Example: $\Gamma = \{a,b,c,d\}$ and $\Delta = \{ \begin{array}{|c|}\hline a \\ \hline b \\ \hline \end{array}, \begin{array}{|c|}\hline b \\ \hline a \\ \hline \end{array}, \begin{array}{|c|}\hline c \\ \hline d \\ \hline \end{array}, \begin{array}{|c|}\hline d \\ \hline c \\ \hline \end{array}, \begin{array}{|c|}\hline x \\ \hline x \\ \hline \end{array}, \begin{array}{|c|c|}\hline a & c \\ \hline \end{array}, \begin{array}{|c|c|}\hline c & a \\ \hline \end{array},$ $\begin{array}{|c|c|}\hline b & d \\ \hline \end{array}, \begin{array}{|c|c|}\hline d & b \\ \hline \end{array}, \begin{array}{|c|c|}\hline x & x \\ \hline \end{array} \mid x \in \Gamma \}$, then

$$\begin{array}{|c|c|}\hline a,b,c,d & a,b,c,d \\ \hline b,d & a,b,c,d \\ \hline \end{array} \underset{sd(\Delta,\pi)}{\Longrightarrow} \begin{array}{|c|c|}\hline a,b,c,d & a,b,c,d \\ \hline b,d & b,d \\ \hline \end{array}$$

Example: $\Gamma = \{a, b, c, d\}$ and $\Delta = \{\begin{array}{|c|}\hline a \\ \hline b \\ \hline \end{array}, \begin{array}{|c|}\hline b \\ \hline a \\ \hline \end{array}, \begin{array}{|c|}\hline c \\ \hline d \\ \hline \end{array}, \begin{array}{|c|}\hline d \\ \hline c \\ \hline \end{array}, \begin{array}{|c|}\hline x \\ \hline x \\ \hline \end{array}, \begin{array}{|c|c|}\hline a & c \\ \hline \end{array}, \begin{array}{|c|c|}\hline c & a \\ \hline \end{array}, \begin{array}{|c|c|}\hline b & d \\ \hline \end{array}, \begin{array}{|c|c|}\hline d & b \\ \hline \end{array}, \begin{array}{|c|c|}\hline x & x \\ \hline \end{array} \mid x \in \Gamma\}$, then

a, b, c, d	c, d
a, b, c, d	a, b, c, d

$$\underset{sd(\Delta,\pi)}{\Longrightarrow}$$

a, b, c, d	c, d
a, b, c, d	c, d

Example: $\Gamma = \{a,b,c,d\}$ and $\Delta = \{\begin{array}{|c|}\hline a \\ \hline b \\ \hline\end{array}, \begin{array}{|c|}\hline b \\ \hline a \\ \hline\end{array}, \begin{array}{|c|}\hline c \\ \hline d \\ \hline\end{array}, \begin{array}{|c|}\hline d \\ \hline c \\ \hline\end{array}, \begin{array}{|c|}\hline x \\ \hline x \\ \hline\end{array}, \begin{array}{|c|c|}\hline a & c \\ \hline\end{array}, \begin{array}{|c|c|}\hline c & a \\ \hline\end{array}, \begin{array}{|c|c|}\hline b & d \\ \hline\end{array}, \begin{array}{|c|c|}\hline d & b \\ \hline\end{array}, \begin{array}{|c|c|}\hline x & x \\ \hline\end{array} \mid x \in \Gamma\}$, then

a,b,c,d	c,d
b,d	a,b,c,d

$$\underset{d(\Delta,\pi)}{\Longrightarrow}$$

a,b,c,d	c,d
b,d	d

Corollary Languages in (S)DREC can be accepted in linear time.

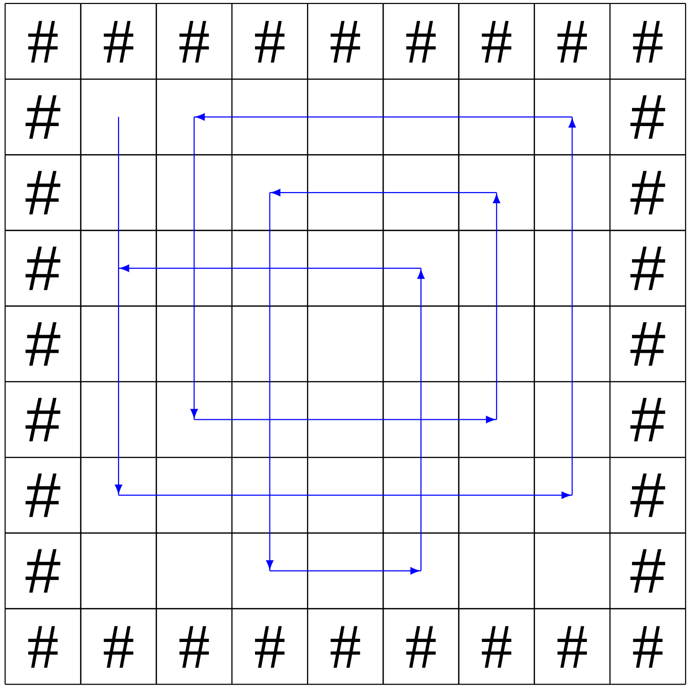
Remark: $\mathscr{L}_d(\Delta, \pi) \subseteq \mathscr{L}_{sd}(\Delta, \pi) \subseteq \mathscr{L}(\Delta, \pi)$.

Theorem[Rei98] The language of pictures over $\{a, b\}$, where all occurring b's are connected is in $MDREC$, $DREC$ and REC.

Theorem[Rei98] $MDREC \subseteq REC$.

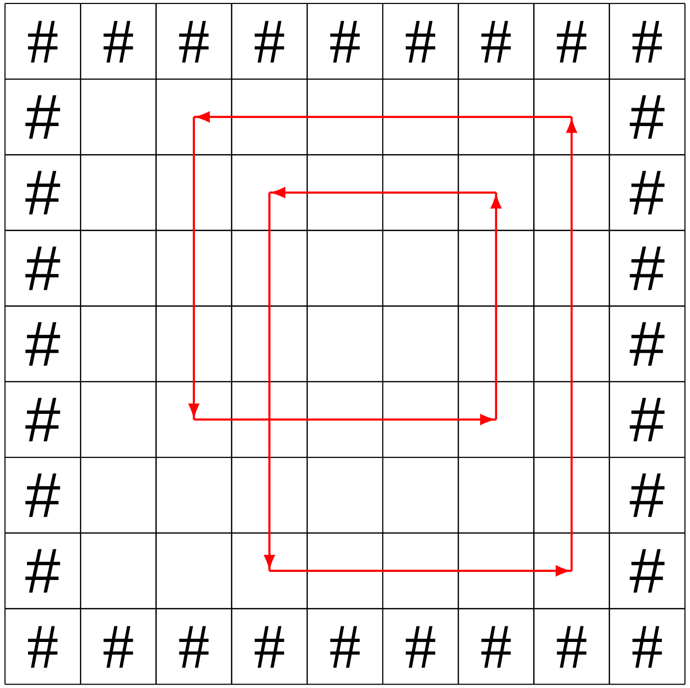
The mirror of a permutation matrix like in [KM01]

#	#	#	#	#	#	#	#	#
#		1		2			1	#
#			1	2		1		#
#	1			2	1			#
#	3	3	3	2	3	3	3	#
#		1		2		1		#
#	1			2			1	#
#			1	2	1			#
#	#	#	#	#	#	#	#	#



Not in *REC* by bisection-argument.

#	#	#	#	#	#	#	#	#
#		1		2			1	#
#			1	2		1		#
#	1			2	1			#
#	3	3	3	2	3	3	3	#
#		1		2		1		#
#	1			2	1			#
#			1	2			1	#
#	#	#	#	#	#	#	#	#



The order by the deterministic process gives more information.

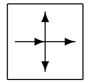
In *SDREC* more general: $4AFA \subseteq SDREC$

Directed acyclic graphs

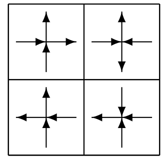
Let $\Gamma = \mathbb{Z}^4$, $\gamma = (l(\gamma), r(\gamma), u(\gamma), d(\gamma)) \in \Gamma$ and $L_{con} := \mathscr{L}(\Delta_{con}) \subseteq \Gamma^{*,*}$

with $\Delta_{con} :=$

$$\left\{ \begin{array}{|c|c|} \hline \gamma & \gamma' \\ \hline \delta & \delta' \\ \hline \end{array} \in (\{\#\} \cup \Gamma)^{2,2} \;\middle|\; \begin{array}{l} (\gamma = (l,r,u,d) \wedge \gamma' = (l',r',u',d')) \rightarrow r = -l', \\ (\delta = (l,r,u,d) \wedge \delta' = (l',r',u',d')) \rightarrow r = -l', \\ (\gamma = (l,r,u,d) \wedge \delta = (l',r',u',d')) \rightarrow d = -u', \\ (\gamma' = (l,r,u,d) \wedge \delta' = (l',r',u',d')) \rightarrow d = -u' \end{array} \right\}$$

Let $\Gamma = \{-1, 1\}^4$ and we identify for example $(-1,1,1,1) =$ .

Now we define an local picture language $L_{locdag} \subseteq L_{con} \cap \Gamma^{*,*}$ where

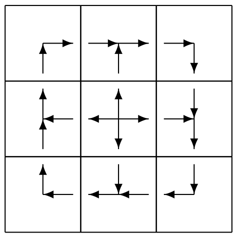
we do not allow sources, sinks and local 4-cycles like i. e. 

$$\Delta_{locdag} := \Delta_{con} \cap (\{\#\} \cup \Gamma \setminus \{(1,1,1,1),(-1,-1,-1,-1)\})^{2,2} \cap$$

$$\left\{ \begin{array}{|c|c|} \hline \gamma & \gamma' \\ \hline \delta & \delta' \\ \hline \end{array} \;\middle|\; \left(\begin{array}{l} \gamma = (l, x_1, u, -x_4) \wedge \gamma' = (-x_1, r, u', x_2) \wedge \\ \delta = (l', -x_3, x_4, d') \wedge \delta' = (x_3, r', -x_2, d) \end{array} \right) \rightarrow \exists i, j \; x_i \neq x_j \right\}$$

Lemma 1 *A picture in* L_{locdag} *describes a directed acyclic graph.*

Let $L_{dag} \subset \{-1,0,1\}^{4^*}$ be the language of pictures describing a directed acyclic graph, $L'_{dag} := L_{dag} \cap \{-1,1\}^{4^*}$,
$L^{-}_{dag} := L_{dag} \cap (\{-1,1\}^4 \setminus \{(1,1,1,1)\})^*$ and
$L^{+}_{dag} := L_{dag} \cap (\{-1,1\}^4 \setminus \{(-1,-1,-1,-1)\})^*$.

Cycles not locally detectable: 

We will now prove the following chain of implications: Lemma 1 $\Rightarrow L^{-}_{dag} \in REC \Rightarrow L'_{dag} \in REC \Rightarrow L_{dag} \in REC \Rightarrow DREC \subseteq REC$

Theorem 1 $DREC \subseteq REC$

Proof: For a given picture $p \in \Sigma^{*,*}$ guess a $p' \in (\Sigma \times \{-1,0,1\}^4)^{*,*}$ such that $p = \pi(p')$ is the projection to the first component and the second component is in L_{dag} using Theorem 2.

Then check if for each position in the picture the symbols on these neighbors from which an edge leads to this position are together sufficient to determine the symbol on the position deterministically.

∎

Theorem 2 $L_{dag} \in REC$

Proof: For a given picture $p \in \{-1, 0, 1\}^{4^{*,*}}$ guess for every occuring 0 either -1 or 1 and check if the resulting p' is in L'_{dag} using Theorem 3. ■

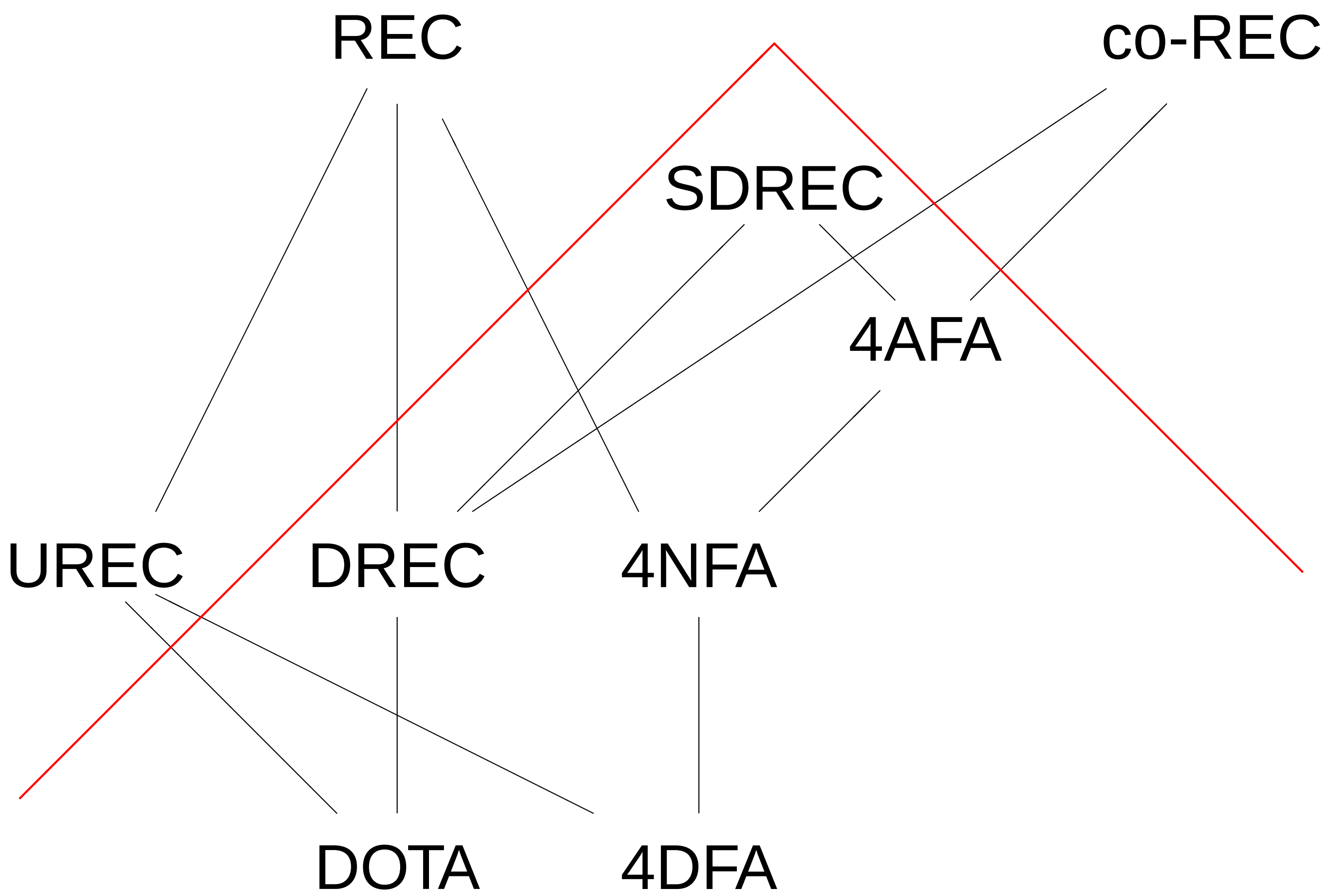
This solves an open problem in [KM01].

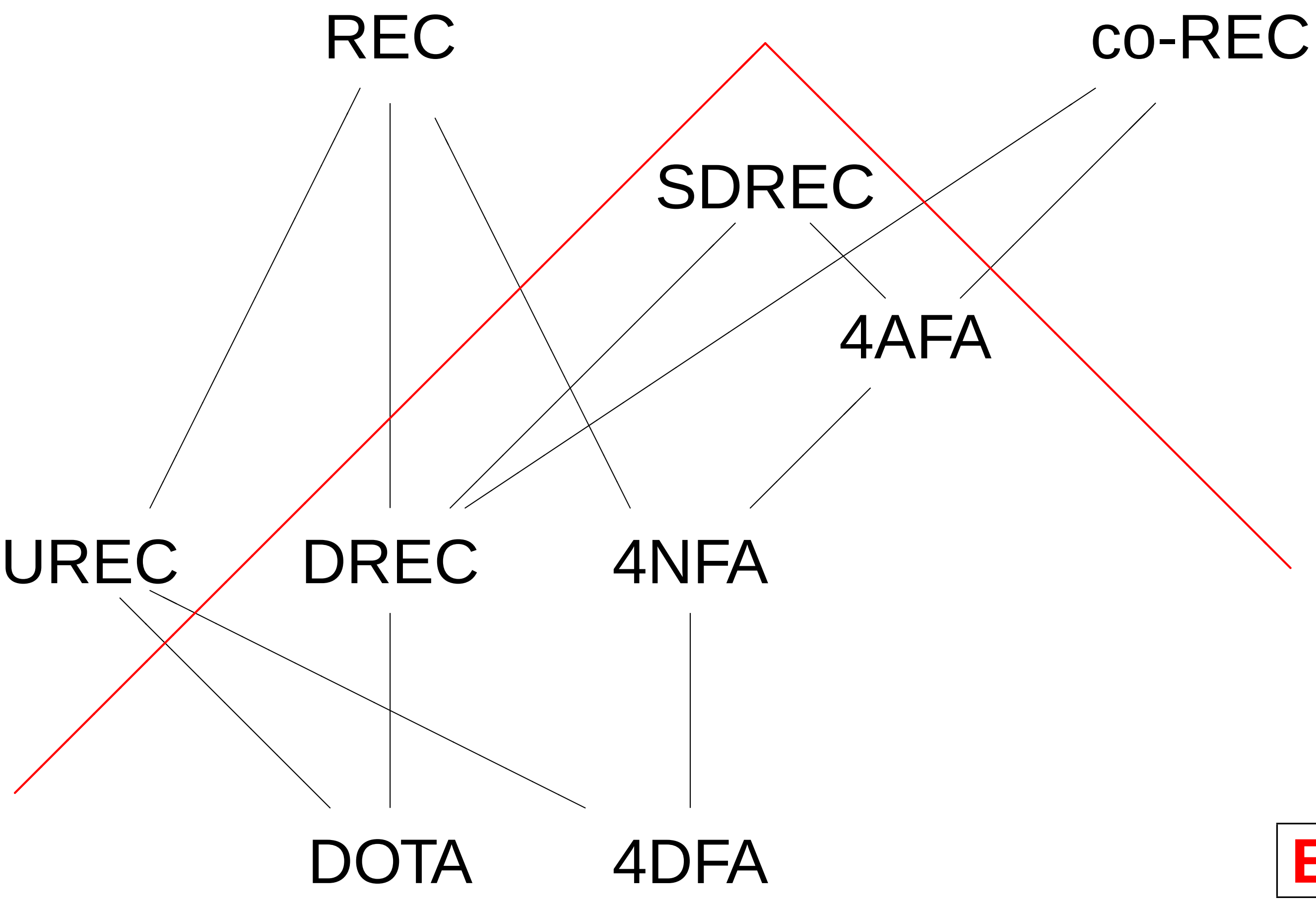
Theorem 3 $L'_{dag} \in REC$

Idea: guess and locally verify a set S of edges where we turn around the direction of arrows obtaining a picture in L^{-}_{dag} without destroying a cycle and apply Theorem 4.

Theorem 4 $L^{-}_{dag}, L^{+}_{dag} \in REC$

Idea: Iterate previous method obtaining a picture in L_{locdag} without destroying a cycle, apply Lemma 1.


REC
co-REC
SDREC
4AFA
UREC
DREC
4NFA
DOTA
4DFA



End