VMCAI 2008

Abstract Interpretation of Reachable Complexes in Biological Signalling Networks

Jérôme Feret

Harvard Medical School

École Normale Supérieure

http://www.di.ens.fr/~feret

Joint work with

Vincent Danos, Walter Fontana, and Jean Krivine

January, 2008

Co-authors

Walter Fontana
Harvard Medical School

Vincent Danos
Paris VII

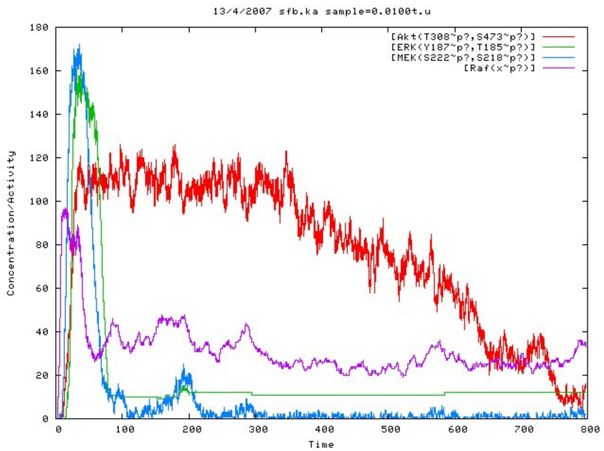
Jean Krivine
École polytechnique

Overview

1. Introduction
2. Kappa language
3. Local views
4. Local set of complexes
5. Local rule systems
6. Conclusion

Modeling signaling pathway

- A cell measures (i.e. checks thresholds, integrates, compares) the concentration of some proteins in order to make decisions.
- Many proteins (enzymes, receptors, transport molecules) are involved. They interact by binding with each other and activating each other.
- We want to track the evolution of some species:



- There is a combinatorial blow-up.

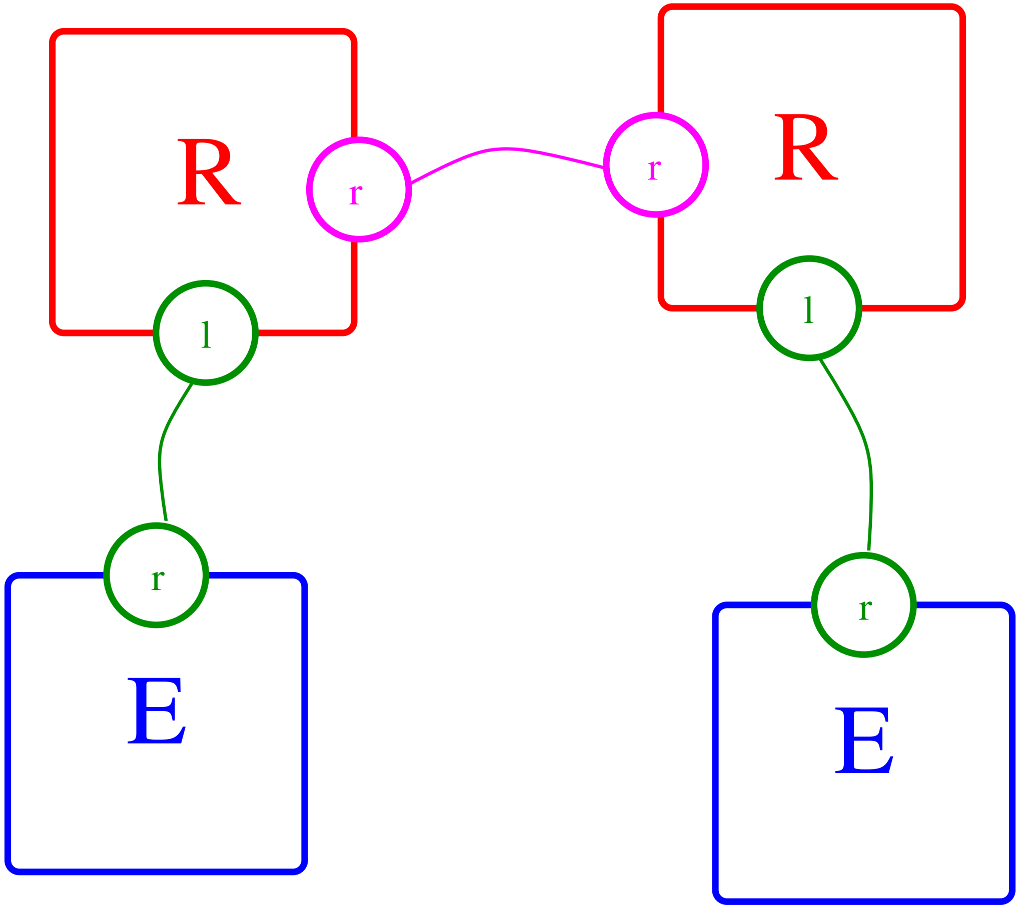
Why using modelling tools ?

- Use a concise high-level description of what happens;
- Share parts of models;
- Derive quantitative models automatically:
  - run benchs of simulations,
  - modify initial conditions,
  - update/modify the model at no cost;
- Use static analysis tools in order to check the consistency of a model:
  - dead rules detection,
  - control detection (which site controls which binding),
  - wrong species detection.

Overview

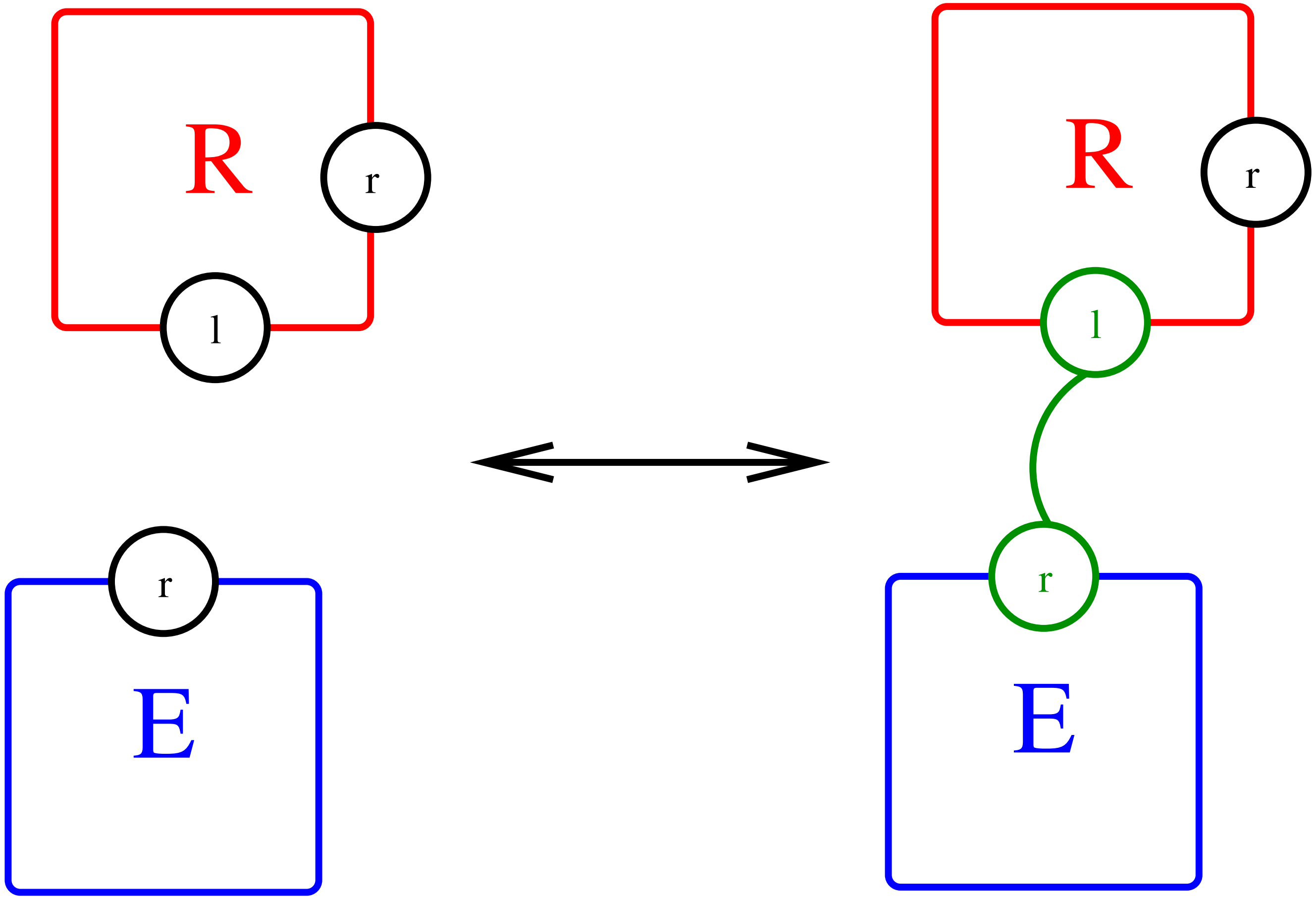
1. Introduction
2. Kappa language
3. Local views
4. Local set of complexes
5. Local rule systems
6. Conclusion

A complex



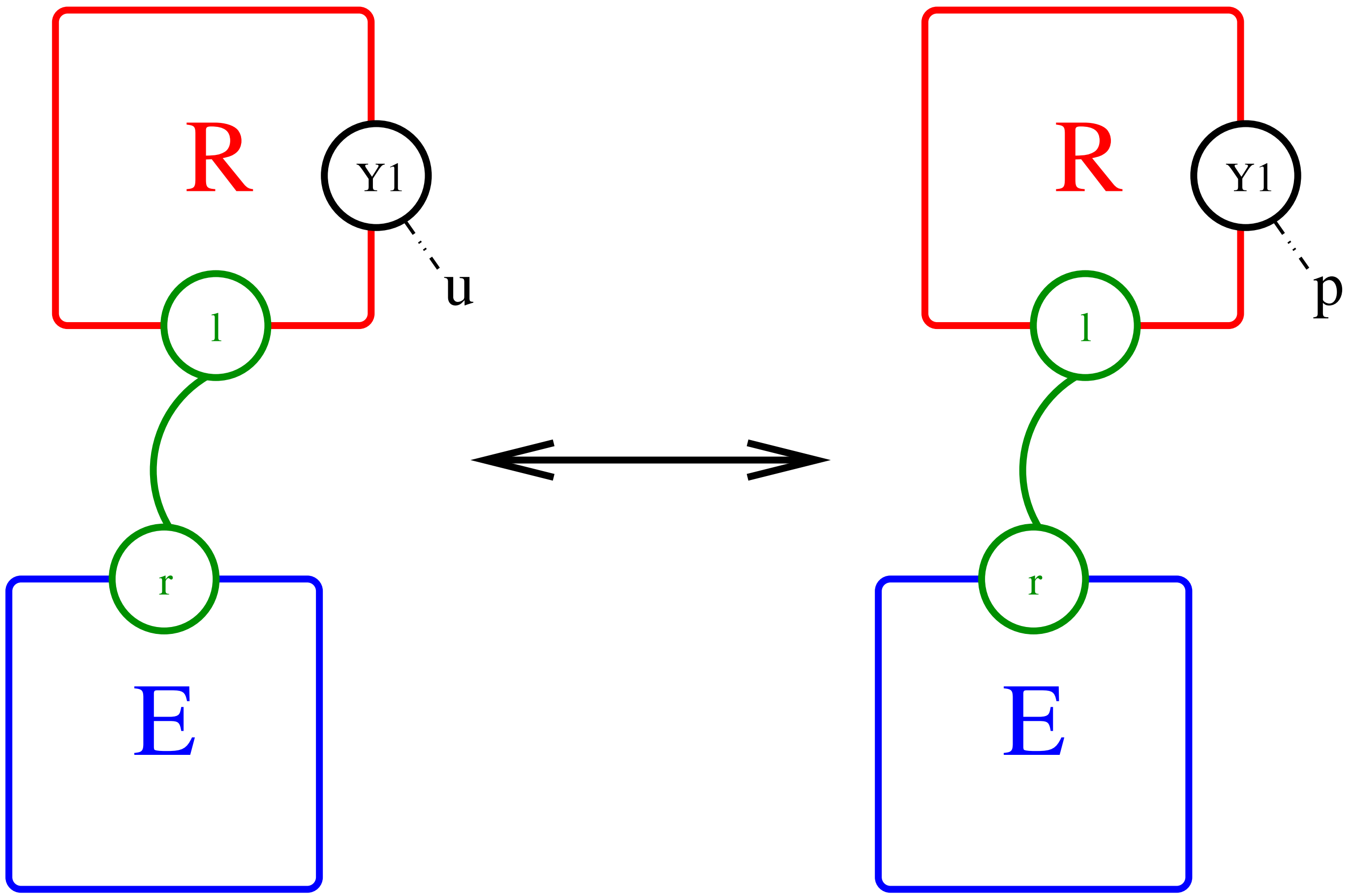
E(r!1), R(l!1,r!2), R(r!2,l!3), E(r!3)

A unbinding/binding Rule



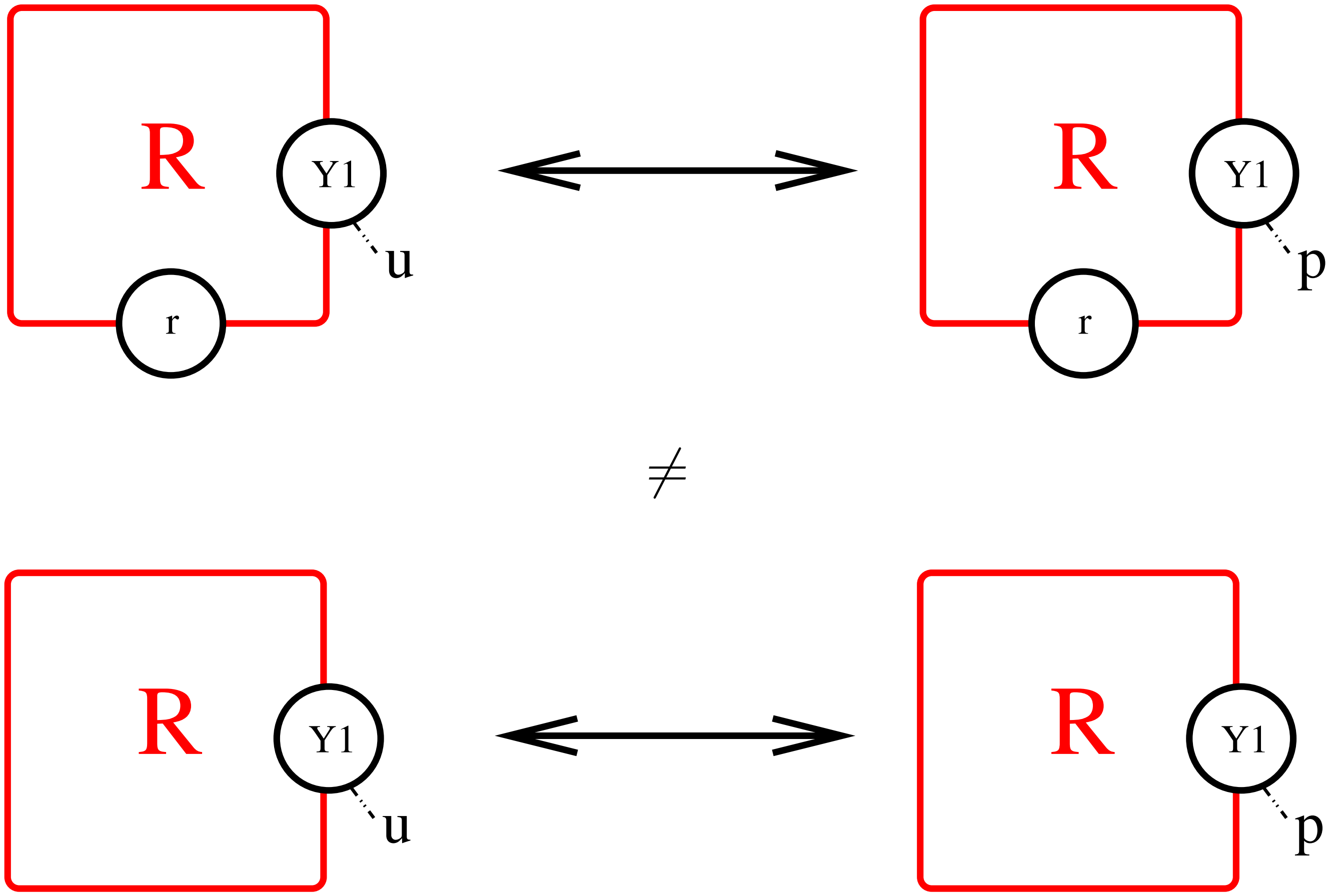
E(r), R(l,r) $\longleftrightarrow$ E(r!1), R(l!1,r)

Internal state

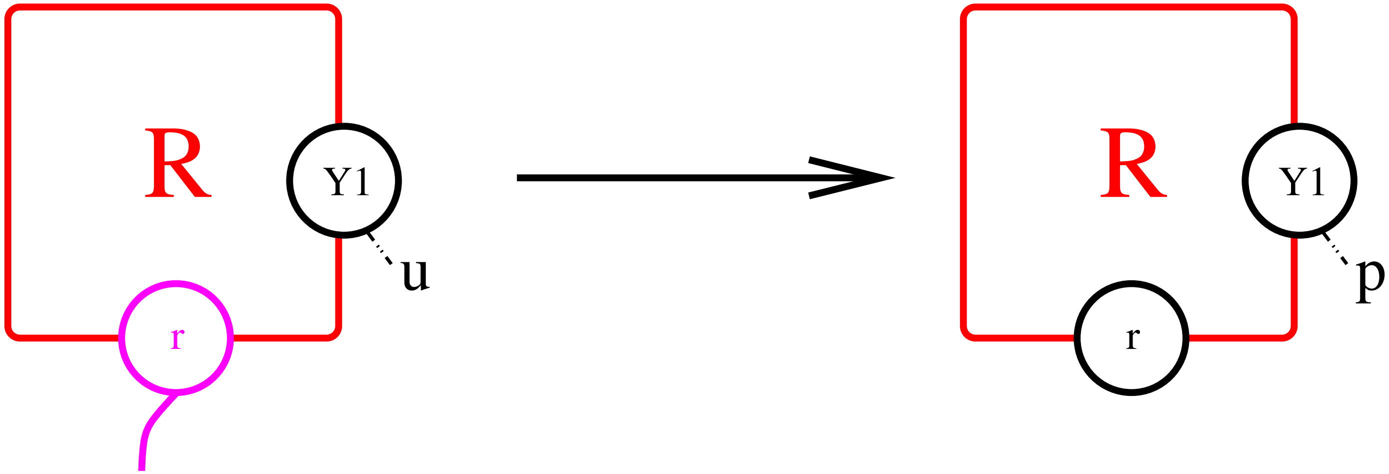


R(Y1~u,l!1), E(r!1) $\longleftrightarrow$ R(Y1~p,l!1), E(r!1)

Don't care, Don't write

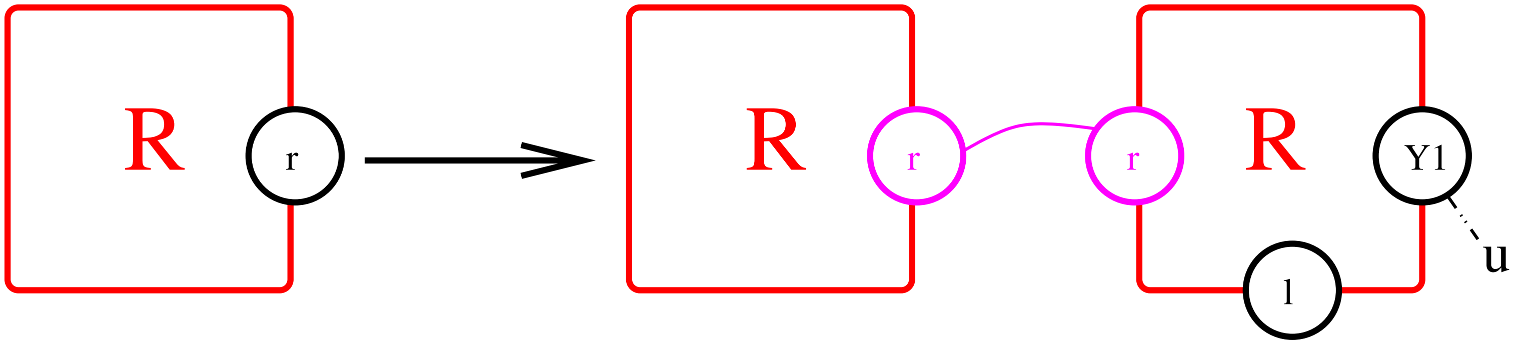


A contextual rule

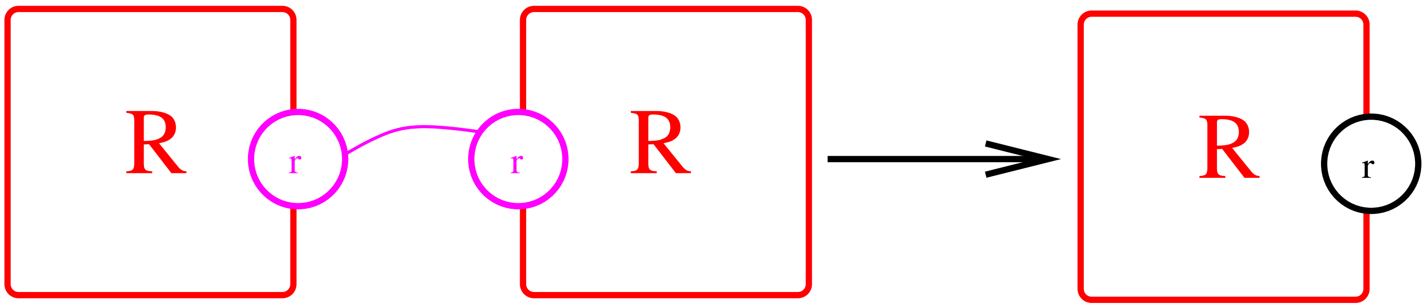


R(Y1~u,r!_) $\rightarrow$ R(Y1~p,r)

Creation/Suppression



R(r) $\rightarrow$ R(r!1), R(r!1,l,Y1)



R(r!1), R(r!1) $\rightarrow$ R(r)

Overview

Set of reachable complexes

Let $\mathcal{R} = \{R_i\}$ be a set of rules.
Let *Complex* be the set of all complexes ($C, c_1, c'_1, \ldots, c_k, c'_k, \ldots \in$ *Complex*).
Let *Complex*$_0$ be the set of initial complexes.
We write:

$$c_1, \ldots, c_m \rightarrow_{R_k} c'_1, \ldots, c'_n$$

whenever:

1. there is an injection of the lhs of R_k in the solution $c_1, \ldots, c_m$;
2. the (injection/rule) produces the solution $c'_1, \ldots, c'_n$.

We are interested in *Complex*$_\omega$ the set of all complexes that can be constructed in one or several applications of rules in $\mathcal{R}$ starting from the set *Complex*$_0$ of initial complexes.

(We do not care about the number of occurrences of each complex).

Inductive definition

We define the mapping $\mathbb{F}$ as follows:

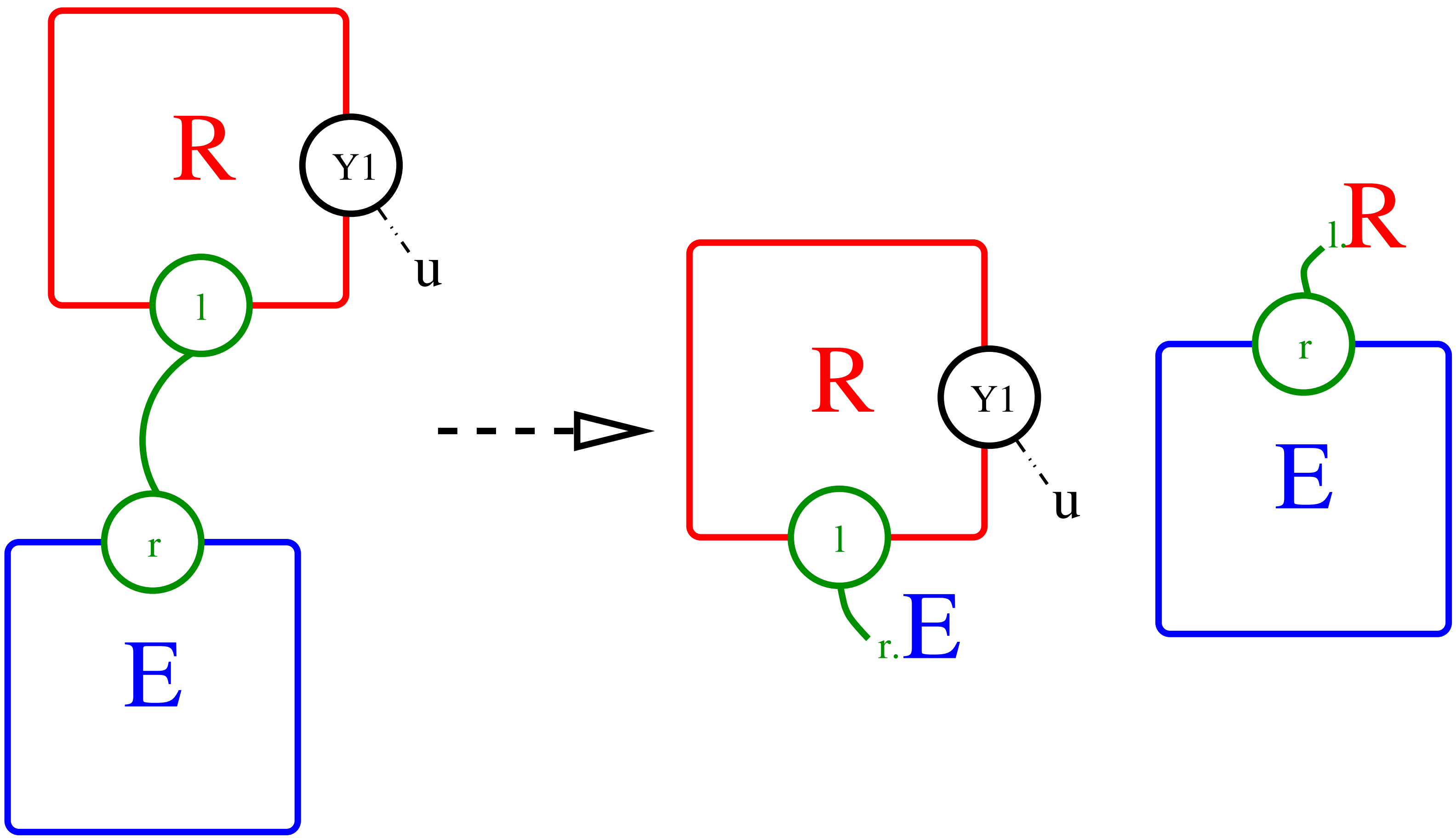
$$\mathbb{F} : \begin{cases} \wp(\mathit{Complex}) & \rightarrow \wp(\mathit{Complex}) \\ X & \mapsto X \cup \left\{ c'_j \;\middle|\; \begin{array}{l} \exists R_k \in \mathcal{R}, c_1, \ldots, c_m \in X, \\ c_1, \ldots, c_m \rightarrow_{R_k} c'_1, \ldots, c'_n \end{array} \right\} \end{cases}.$$

The set $\wp(\mathit{Complex})$ is a complete lattice.
The mapping $\mathbb{F}$ is an extensive $\cup$-complete morphism.

We have:

$$\mathit{Complex}_\omega = \bigcup \left\{ \mathbb{F}^n(\mathit{Complex}_0) \;\middle|\; n \in \mathbb{N} \right\}.$$

Local views



$\alpha(\{R(Y1{\sim}u, l!1), E(r!1)\}) = \{R(Y1{\sim}u, l!r.E); E(r!l.R)\}.$

Galois connexion

Let *Local_view* be the set of all local views.

Let $\alpha \in \wp(\mathit{Complex}) \rightarrow \wp(\mathit{Local\_view})$ be the function that maps any set of complexes into the set of their local views.

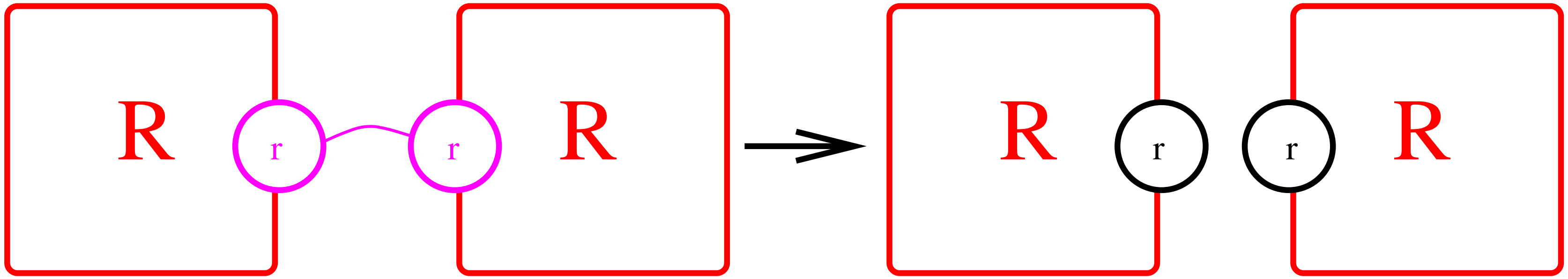
The set $\wp(\mathit{Local\_view})$ is a complete lattice.
The function α is a $\cup$-complete morphism.

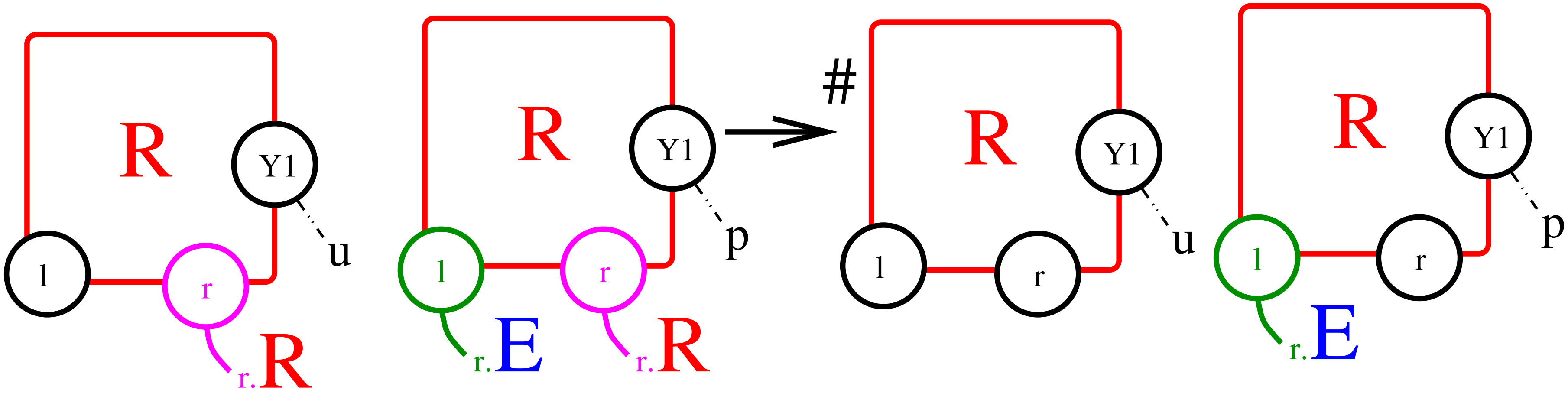
Thus, it defines a Galois connexion:

$$\wp(\mathit{Complex}) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} \wp(\mathit{Local\_view}).$$

(The function γ maps a set of local views into the set of complexes that can be built with these local views).

Abstract rules





Abstract counterpart to $\mathbb{F}$

We define $\mathbb{F}^\sharp$ as:

$$\mathbb{F}^\sharp : \begin{cases} \wp(\mathit{Local\_view}) & \to \wp(\mathit{Local\_view}) \\ X & \mapsto X \cup \left\{ \mathit{lv}'_j \;\middle|\; \begin{array}{l} \exists R_k \in \mathcal{R}, \mathit{lv}_1, \ldots, \mathit{lv}_m \in X, \\ \mathit{lv}_1, \ldots, \mathit{lv}_m \to^\sharp_{R_k} \mathit{lv}'_1, \ldots, \mathit{lv}'_n \end{array} \right\}. \end{cases}$$

Soundness

We have:

1. $(\wp(\mathit{Complex}), \subseteq, \cup)$ and $(\wp(\mathit{Local\_view}), \subseteq, \cup)$ are chain-complete partial orders;
2. $\wp(\mathit{Complex}) \xrightleftharpoons[\alpha]{\gamma} \wp(\mathit{Local\_view})$ is a Galois connexion;
3. $\mathbb{F} \in \wp(\mathit{Complex}) \to \wp(\mathit{Complex})$ and $\mathbb{F}^\sharp \in \wp(\mathit{Local\_view}) \to \wp(\mathit{Local\_view})$ are extensive and monotonic mappings;
4. $\mathbb{F} \circ \gamma \mathrel{\dot{\subseteq}} \gamma \circ \mathbb{F}^\sharp$;

So:

1. both $\mathit{lfp}_{x_0}\mathbb{F}$ and $\mathit{lfp}_{\alpha(x_0)}\mathbb{F}^\sharp$ exist,
2. $\mathit{lfp}_{x_0}\mathbb{F} \subseteq \gamma(\mathit{lfp}_{\alpha(x_0)}\mathbb{F}^\sharp)$.

Overview

1. Introduction
2. Kappa language
3. Local views
4. Local set of complexes
5. Local rule systems
6. Conclusion

Which information is abstracted away ?

Our analysis is exact (no false positive):

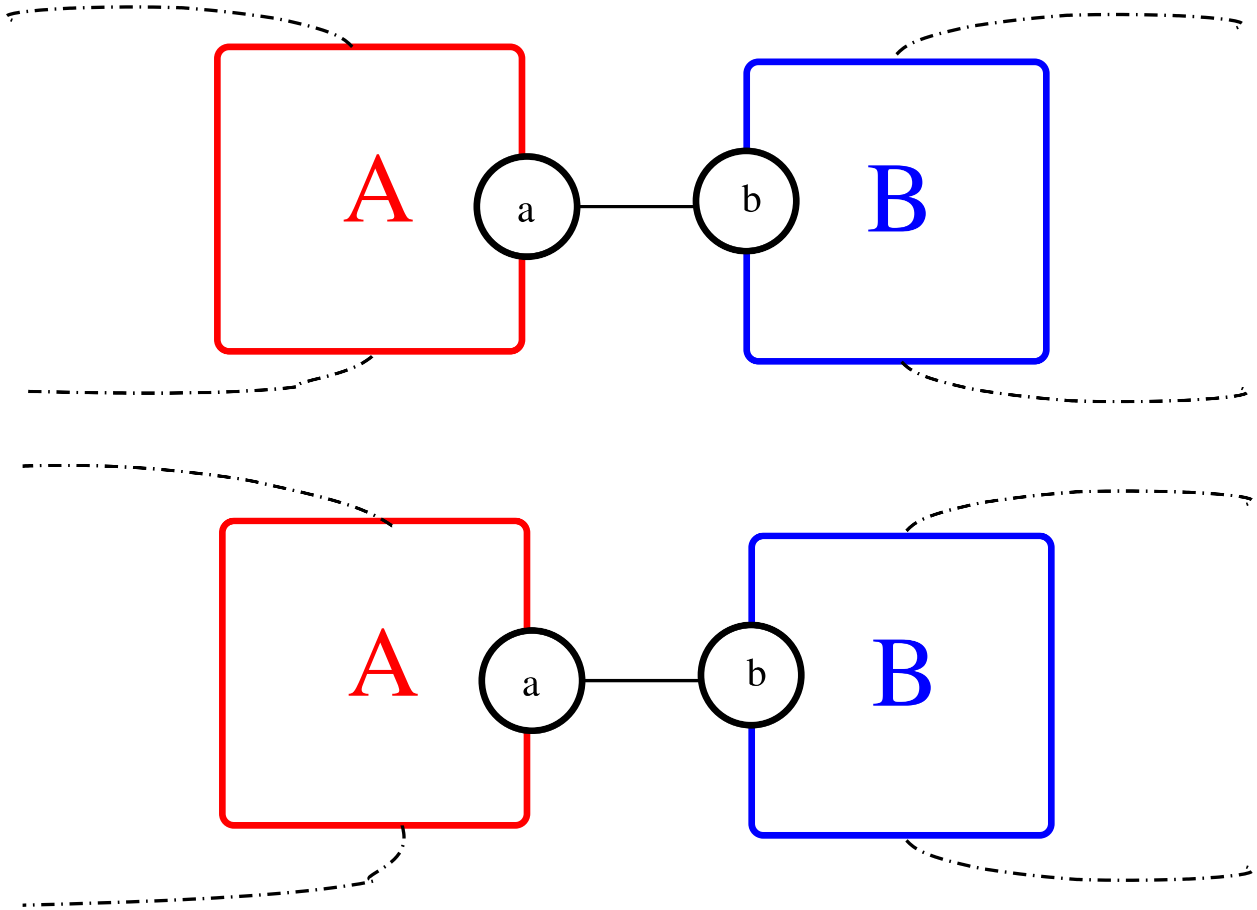
- for the early EGF cascade (356 complexes);
- for the early FGF cascade ($709\,698$ complexes);
- for the EGF cascade with RAS-ERK activation ($\simeq 1.8 * 10^{18}$ complexes);

We know how to build systems with false positives. . .
. . .but they seem to be biologically meaningless.

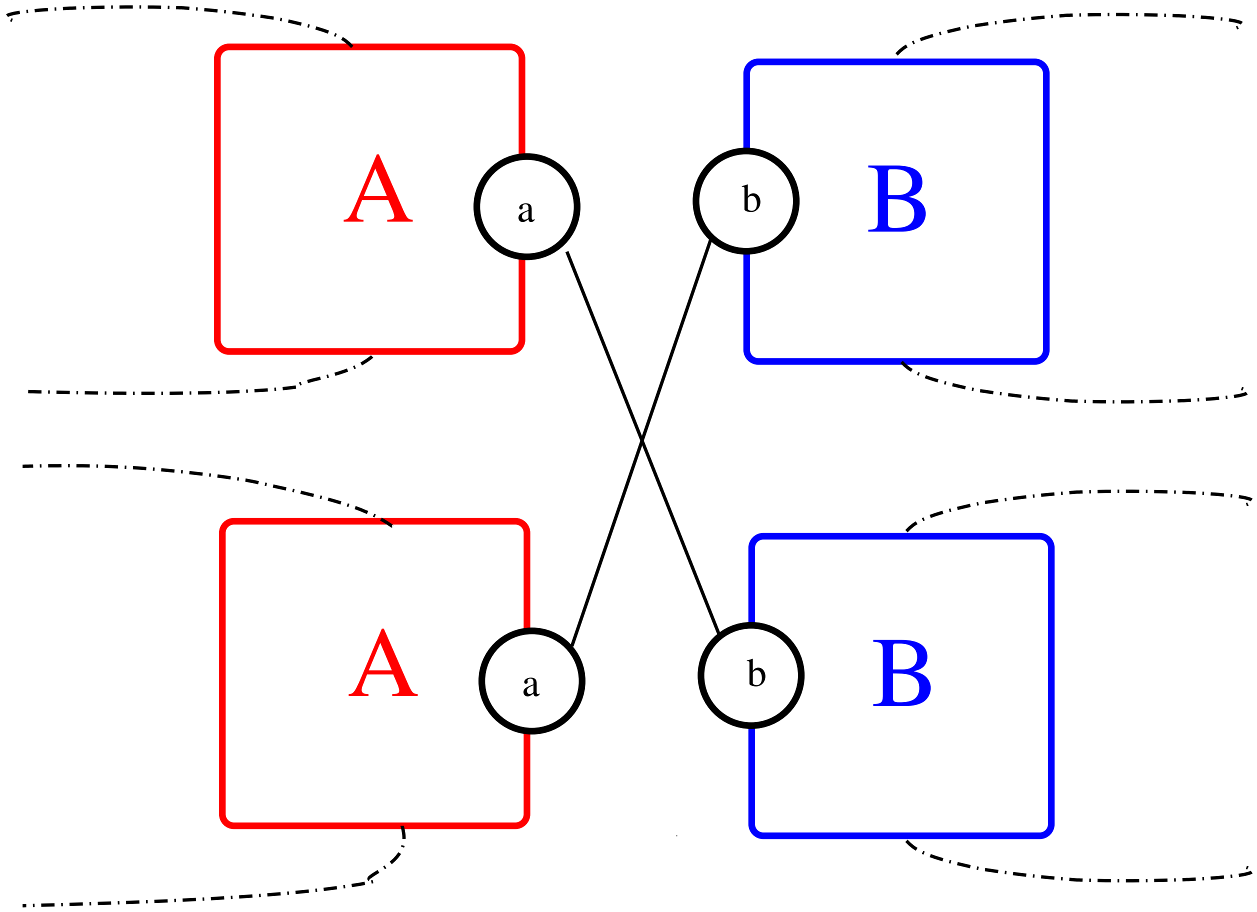
This raises the following issues:

- Can we characterize which information is abstracted away ?
- Which is the form of the systems, for which we have no false positive ?
- Do we learn something about the biological systems that we describe ?

Swap-closure



Swap-closure



Local set of complexes

Let $X \subseteq$ *Complex* be a set of complexes,

The following assertions are equivalent:

1. $X = \gamma(\alpha(X))$
2. X is closed upon swap.

In such a case, we say that X is local.

As a consequence:

1. Any assembling of views in $\alpha(X)$ can be extended in a close complex in $\gamma(\alpha(X))$;
2. We have $\mathbb{F}^\sharp \circ \alpha = \alpha \circ \mathbb{F} \circ \gamma \circ \alpha$.

When is there no false positive ?

We have:

1. $(\wp(\mathit{Complex}), \subseteq, \cup)$ and $(\wp(\mathit{Local\_view}), \subseteq, \cup)$ are chain-complete partial orders;
2. $(\wp(\mathit{Complex}), \subseteq) \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} (\wp(\mathit{Local\_view}), \subseteq)$ is a Galois connexion;
3. $\mathbb{F} : \wp(\mathit{Complex}) \to \wp(\mathit{Complex})$ is an extensive and monotonic map;
4. $\mathbb{F}^\sharp \circ \alpha = \alpha \circ \mathbb{F} \circ \gamma \circ \alpha$

So:

$$\mathit{lfp}_{x_0}\mathbb{F} \in \gamma(\wp(\mathit{Complex})) \iff \mathit{lfp}_{x_0}\mathbb{F} = \gamma(\mathit{lfp}_{\alpha(x_0)}\mathbb{F}^\sharp).$$

Overview

1. Introduction
2. Kappa language
3. Local views
4. Local set of complexes
5. Local rule systems
6. Conclusion

Local fragment of Kappa

If:

1. initial agents are not bound;
2. rules are atomic;
3. rules are local (only agents that interact are tested);
4. binding rules do not interfere i.e. if both:
   - A(a~m,S),B(b~n,T) $\rightarrow$ A(a~m!1,S),B(b~n!1,T)
   - and A(a~m',S'),B(b~n',T') $\rightarrow$ A(a~m'!1,S'),B(b~n'!1,T'),

   then:
   - A(a~m,S),B(b~n',T') $\rightarrow$ A(a~m!1,S),B(b~n'!1,T');
5. assembled complexes are acyclic;

then:

$$Complex_\omega = \gamma(\alpha(Complex_\omega)).$$

Non local systems

$$Complex_0 \triangleq \text{R(a\textasciitilde u)}$$

$$Rules \triangleq \left\{ \begin{array}{lcl} \text{R(a\textasciitilde u)} & \leftrightarrow & \text{R(a\textasciitilde p)} \\ \text{R(a\textasciitilde u),R(a\textasciitilde u)} & \rightarrow & \text{R(a\textasciitilde u!1),R(a\textasciitilde u!1)} \\ \text{R(a\textasciitilde p),R(a\textasciitilde u)} & \rightarrow & \text{R(a\textasciitilde p!1),R(a\textasciitilde p!1)} \\ \text{R(a\textasciitilde p),R(a\textasciitilde p)} & \rightarrow & \text{R(a\textasciitilde p!1),R(a\textasciitilde p!1)} \end{array} \right\}$$

R(a~u!1),R(a~u!1) $\in$ $Complex_\omega$
R(a~p!1),R(a~p!1) $\in$ $Complex_\omega$
But R(a~u!1),R(a~p!1) $\notin$ $Complex_\omega$.

Non local systems

$$\mathit{Complex}_0 \stackrel{\Delta}{=} \text{A(a\textasciitilde u),B(a\textasciitilde u)}$$

$$\mathit{Rules} \stackrel{\Delta}{=} \left\{ \begin{array}{l} \text{A(a}{\sim}\text{u),B(a}{\sim}\text{u)} \rightarrow \text{A(a}{\sim}\text{u!1),B(a}{\sim}\text{u!1)} \\ \text{A(a}{\sim}\text{u!1),B(a}{\sim}\text{u!1)} \rightarrow \text{A(a}{\sim}\text{p!1),B(a}{\sim}\text{u!1)} \\ \text{A(a}{\sim}\text{u!1),B(a}{\sim}\text{u!1)} \rightarrow \text{A(a}{\sim}\text{u!1),B(a}{\sim}\text{p!1)} \end{array} \right\}$$

A(a~u!1),B(a~p!1) $\in \mathit{Complex}_\omega$

A(a~p!1),B(a~u!1) $\in \mathit{Complex}_\omega$

But A(a~p!1),B(a~p!1) $\notin \mathit{Complex}_\omega$.

Program transformation

- we have a syntactic criterion in order to ensure that the set of reachable complexes of a kappa system is local;
- we use program transformations to help systems satisfying this criterion;
  1. decontextualization
     - is fully automatic;
     - preserves the transition system;
     - simplifies rules thanks to reachability analysis.
  2. conjugation
     - manual;
     - preserves the set of reachable complexes;
     - add some rules that are in the transitive closure of the system.

Example

Initial rule:

R(l!2,r),R(l!1,r),E(r!1),E(r!2) $\rightarrow$ R(l!3,r!1),R(l!2,r!1),E(r!2),E(r!3)

Decontextualized rule:

R(l!_,r),R(l!_,r) $\rightarrow$ R(l!_,r!1),R(l!_,r!1)

We can remove redundant tests.

Example

Initial rules:

$$
\begin{array}{rcl}
\text{Sh(Y7\textasciitilde p!2,pi!1),G(a!2,b),R(Y48\textasciitilde p!1)} & \rightarrow & \text{Sh(Y7\textasciitilde p,pi!1),G(a,b),R(Y48\textasciitilde p!1)} \\
\text{Sh(Y7\textasciitilde p!3,pi!1),G(a!3,b!2),So(d!2),R(Y48\textasciitilde p!1)} & \rightarrow & \text{Sh(Y7\textasciitilde p,pi!1),G(a,b!2),So(d!2),R(Y48\textasciitilde p!1)} \\
\text{Sh(Y7\textasciitilde p!1,pi),G(a!1,b)} & \rightarrow & \text{Sh(Y7\textasciitilde p,pi),G(a,b)} \\
\text{Sh(Y7\textasciitilde p!1,pi),G(a!1,b!\_)} & \rightarrow & \text{Sh(Y7\textasciitilde p,pi),G(a,b!\_)}
\end{array}
$$

Decontextualized rule:

Sh(Y7!1),G(a!1) $\rightarrow$ Sh(Y7),G(a)

We can remove exhaustive enumerations.

Overview

1. Introduction
2. Kappa language
3. Local views
4. Local set of complexes
5. Local rule systems
6. Conclusion

Conclusion

- A scalable static analysis to abstract the reachable complexes.
- A class of models for which the abstraction is complete.
- Many applications:
  - idiomatic description of reachable complexes;
  - dead rule detection;
  - rule decontextualization;
  - computer-driven kinetic refinement.
- It can also help simulation algorithms:
  - wake up/inhibition map (agent-based simulation);
  - flat rule system generation (for bounded set of complexes);
  - on the fly flat rule generation (for large/unbounded set).

Future works

- Refine the analysis (and completeness criteria).
  - Deal with locations.
  - Deal with cycles.
  - Shape analysis.
- Quantitative analysis.
  - Can we lift the local-view abstraction to stochastic/differential semantics ?
  - Which information do we obtain ?
- Semi-quantitative abstractions.
  - Can we design abstract domains to discover semi-quantitative properties (i.e. overshoot detection) ?