Reachability Analysis of Rule-based Models

Jérôme Feret

DI - ÉNS

Friday, the 2nd of December, 2016

In this talk...

We illustrate the following concepts:

- Galois connections:
  - the upper closure operator $\gamma \circ \alpha$,
  - the lower closure operator $\alpha \circ \gamma$;

- soundness:
  - the abstraction forgets no behavior;

- completeness:
  - sufficient conditions that ensure the absence of false positive;

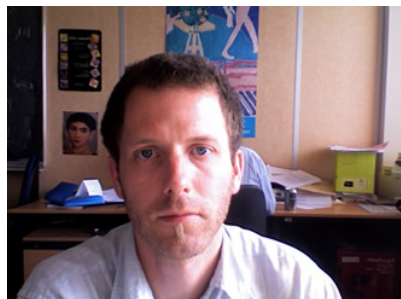on an abstraction of the reachable connected components in a site-graph rewriting language.

Joint-work with...

Walter Fontana
Harvard Medical School
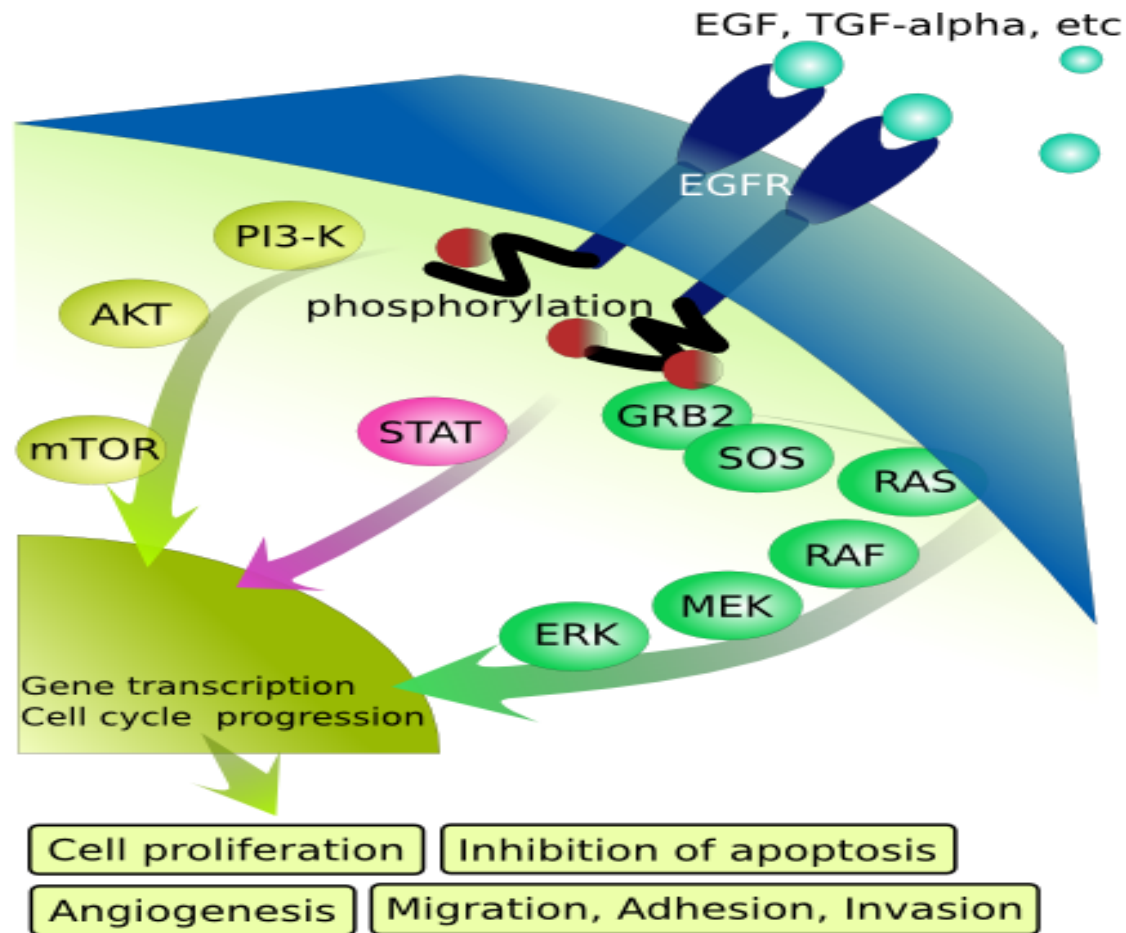
Vincent Danos
ÉNS

Russ Harmer
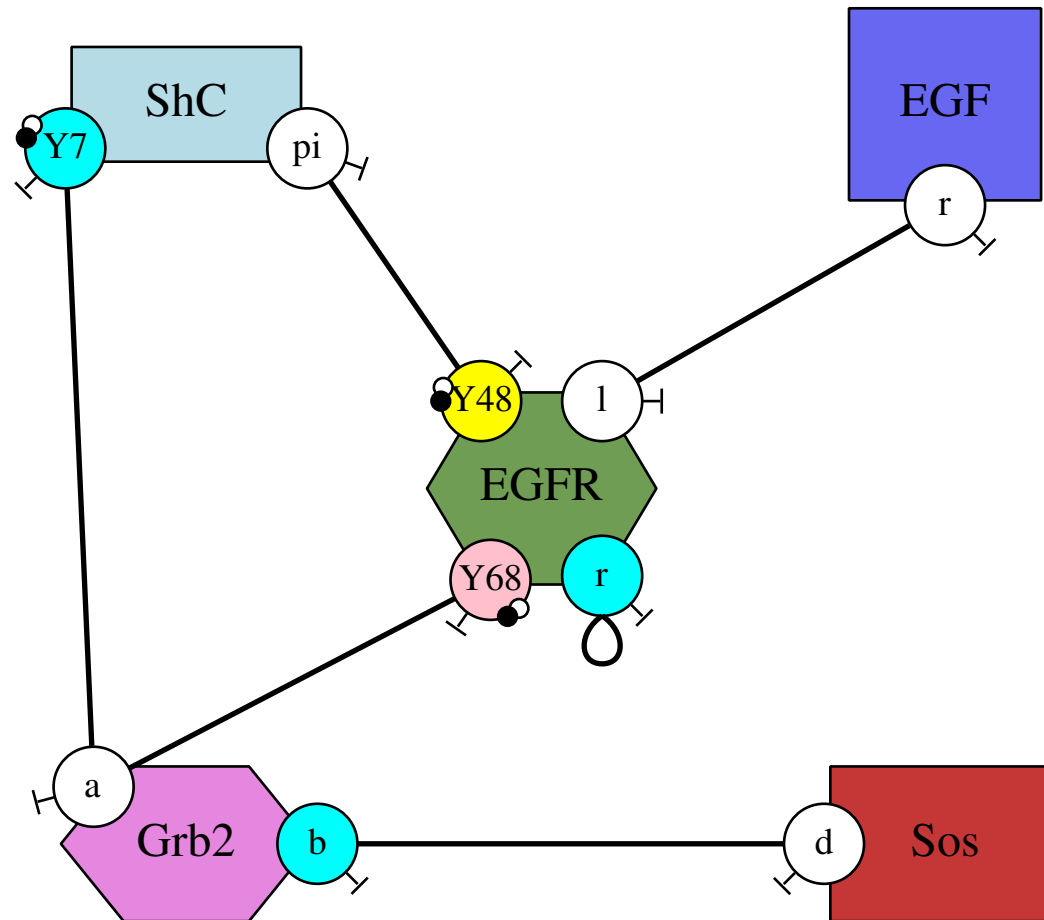ÉNS Lyon

Jean Krivine
Paris VII

Overview

1. Introduction

2. Language: Kappa

3. Abstraction: Local views

4. Completeness: false positives?

5. Local fragment of Kappa

6. Decontextualization

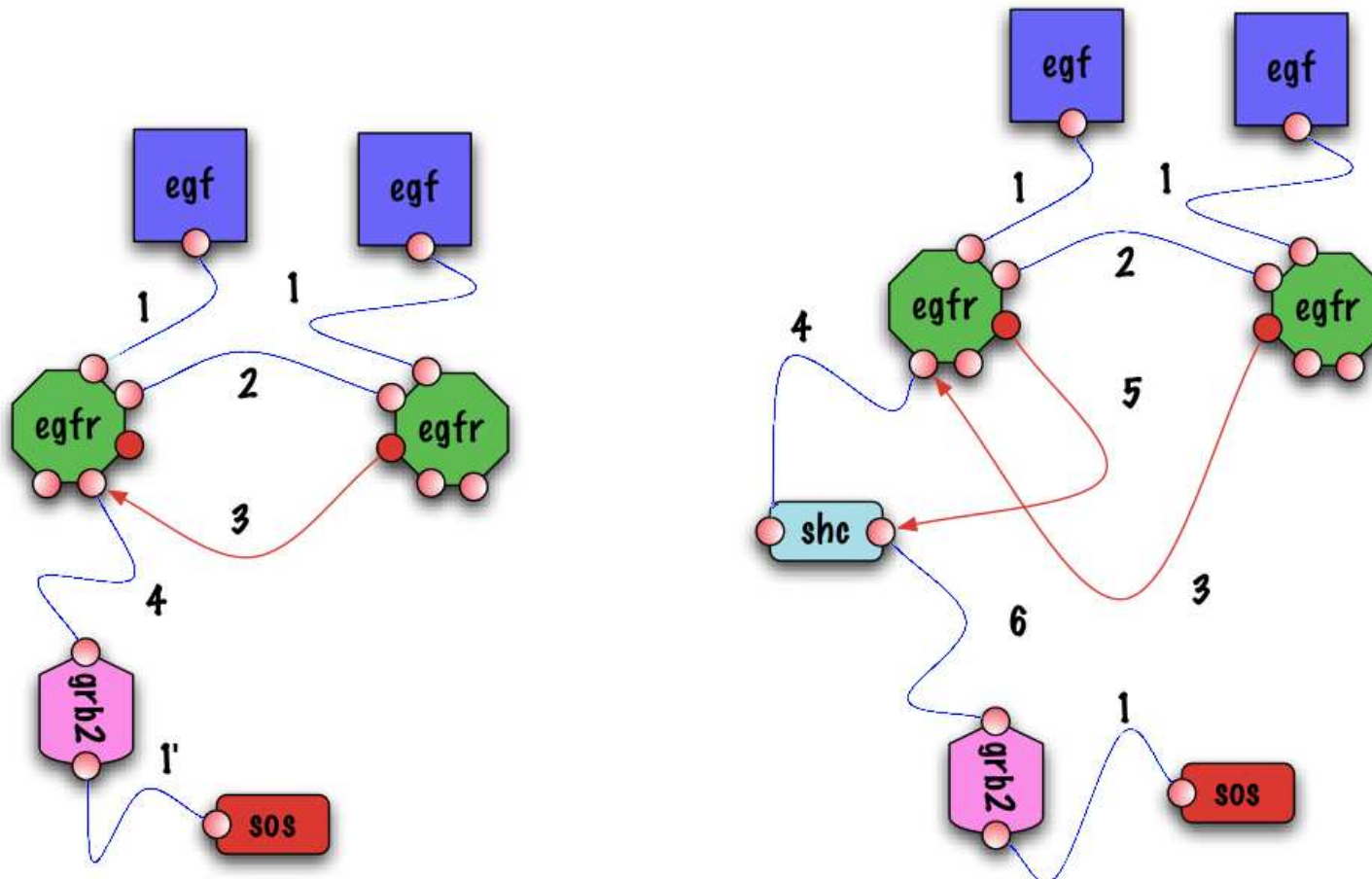7. Conclusion

Signaling Pathways

Eikuch, 2007

Contact map

Causal traces

ODE semantics

EGF pathway (reduced ODEs)

ODE semantics

EGF pathway (reduced ODEs)

What will happen if more Shc(s) is put in the system?

ODE semantics
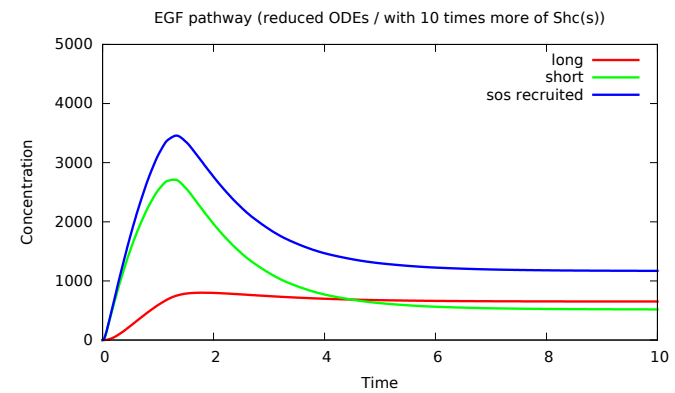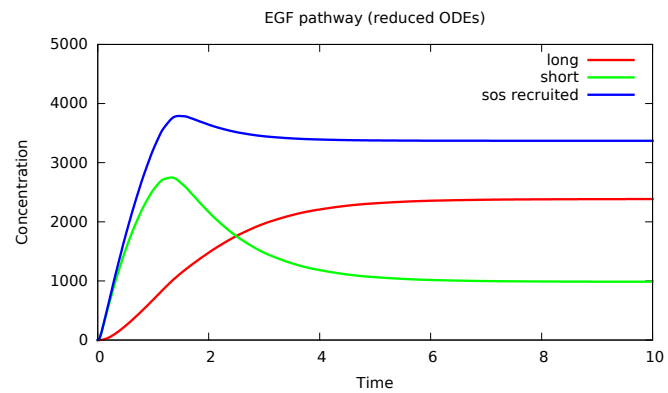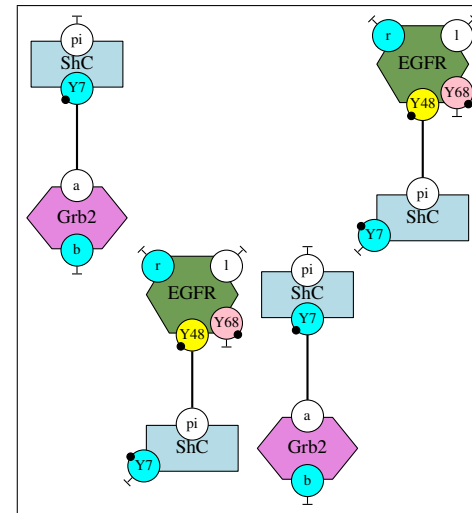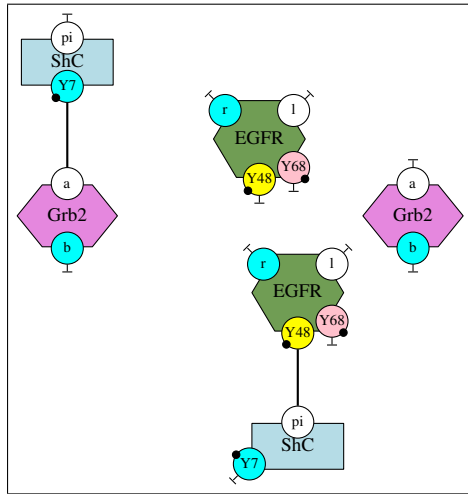
EGF pathway (reduced ODEs)

EGF pathway (reduced ODEs / with 10 times more of Shc(s))

Crowding effect

Overview

1. Introduction

2. Language: Kappa

3. Abstraction: Local views

4. Completeness: false positives?

5. Local fragment of Kappa
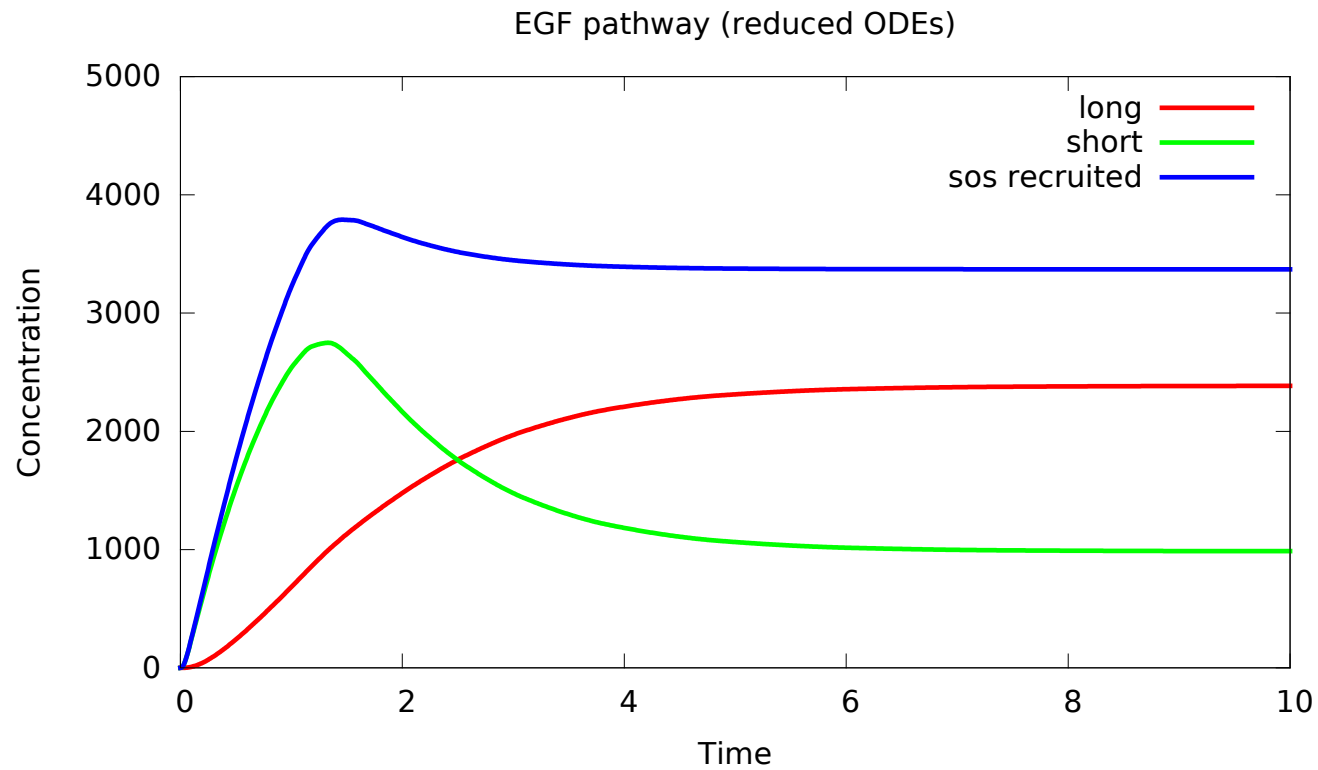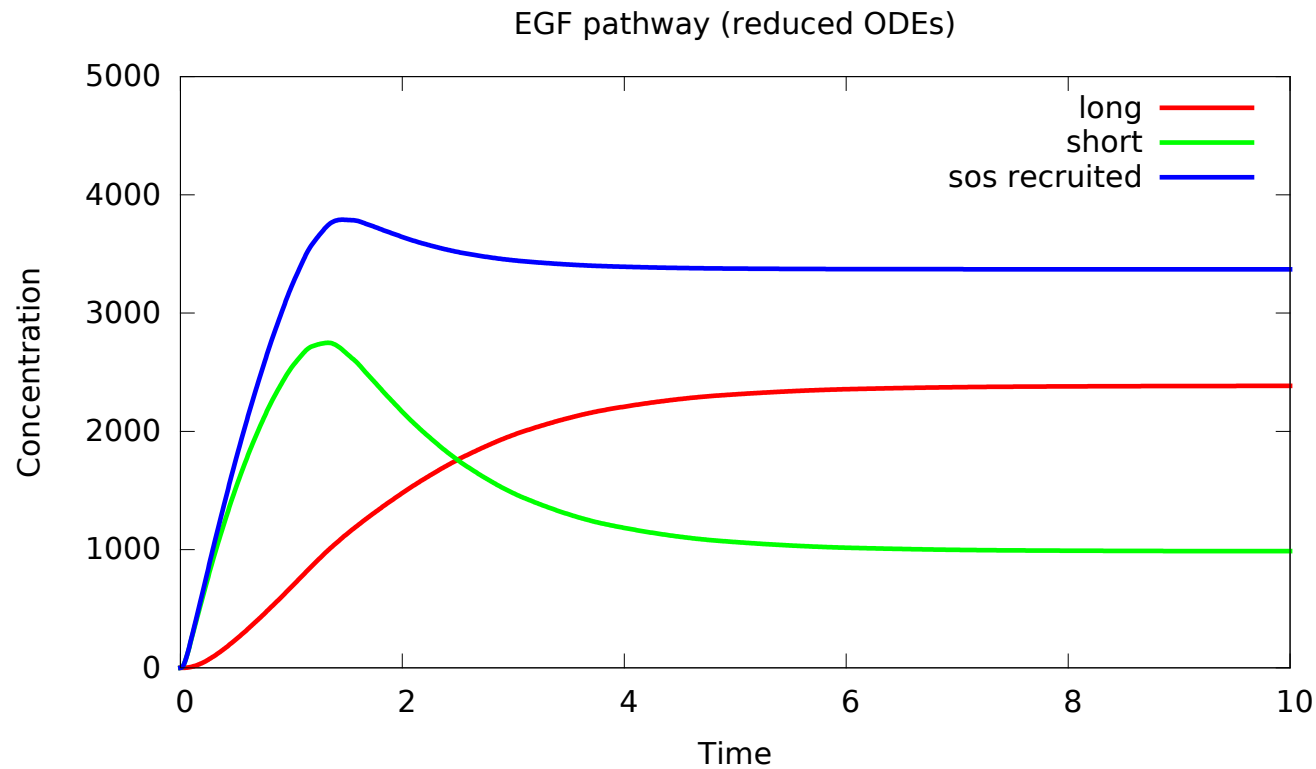
6. Decontextualization

7. Conclusion

A chemical species
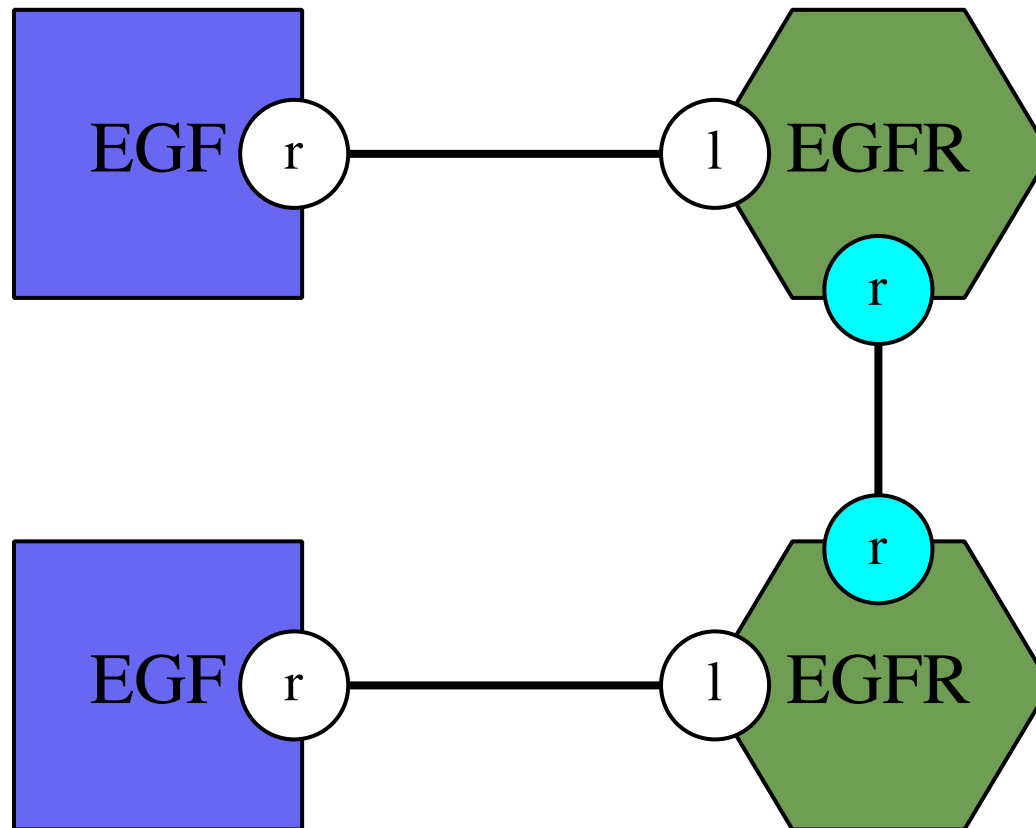
EGF(r!1), EGFR(l!1,r!2), EGFR(r!2,l!3), EGF(r!3)

A Unbinding/Binding Rule
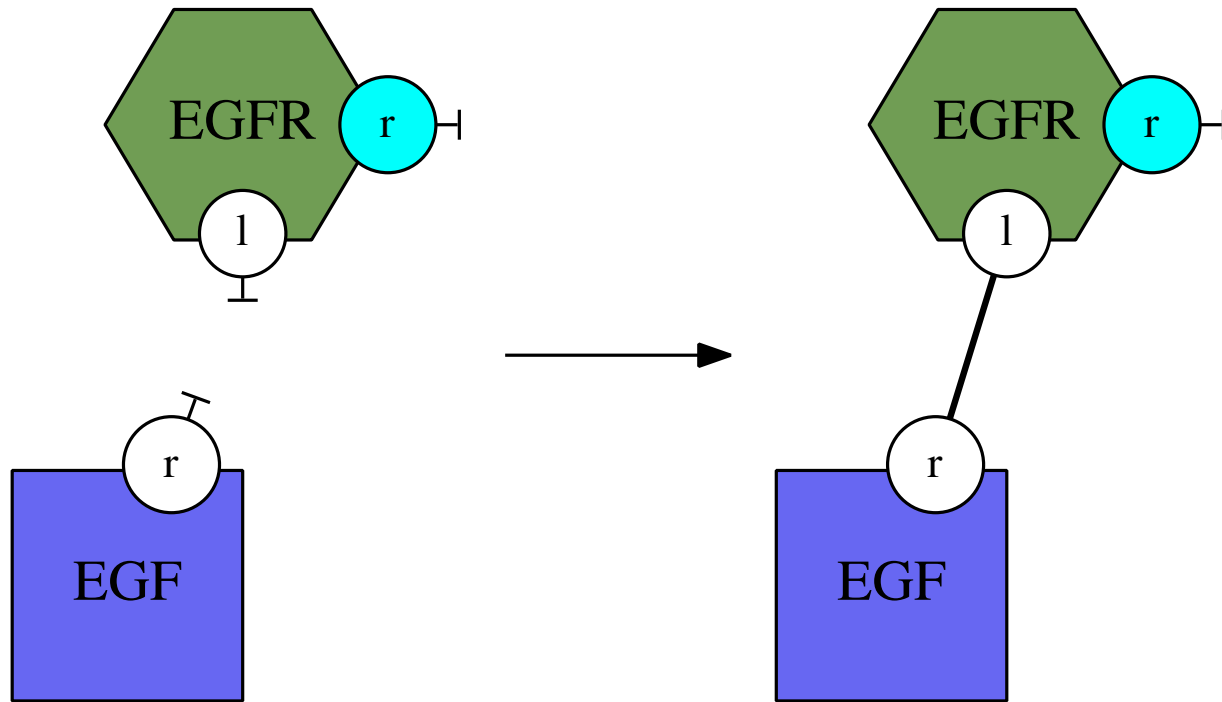
EGF(r), EGFR(l,r) ⟷ EGF(r!1), EGFR(l!1,r)

Internal state
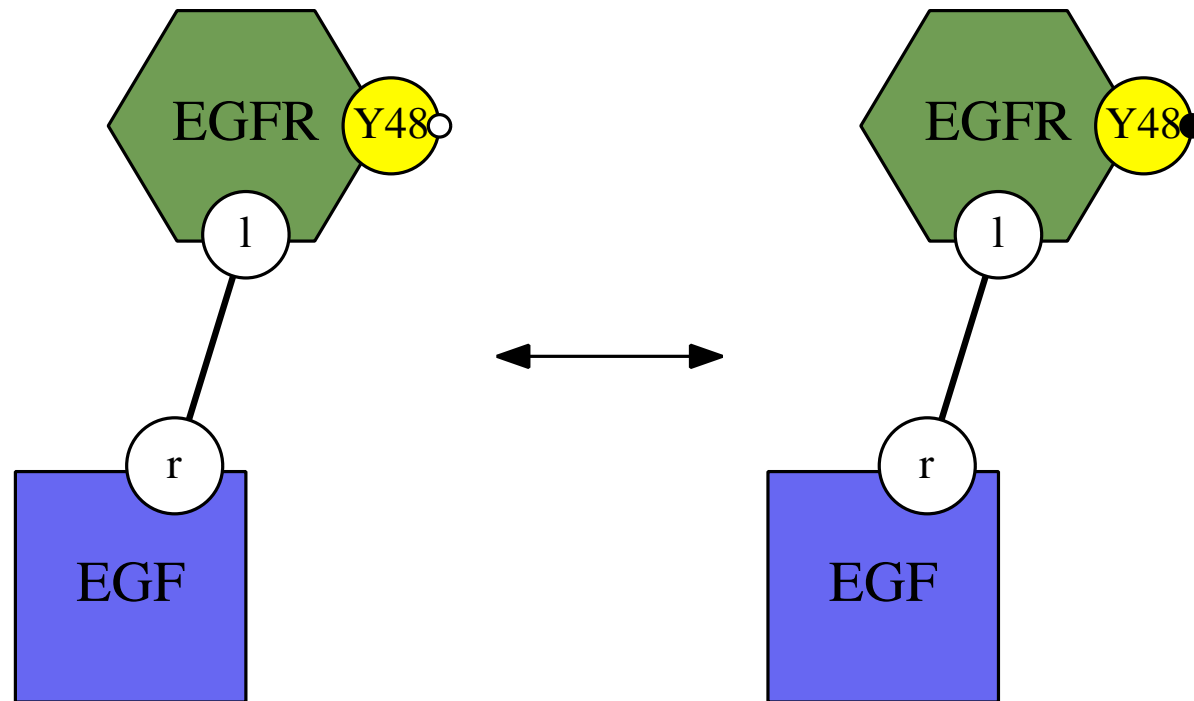
EGFR(Y48~u?,l!1), EGF(r!1) ⟷ EGFR(Y48~p?,l!1), EGF(r!1)

Don't care, Don't write
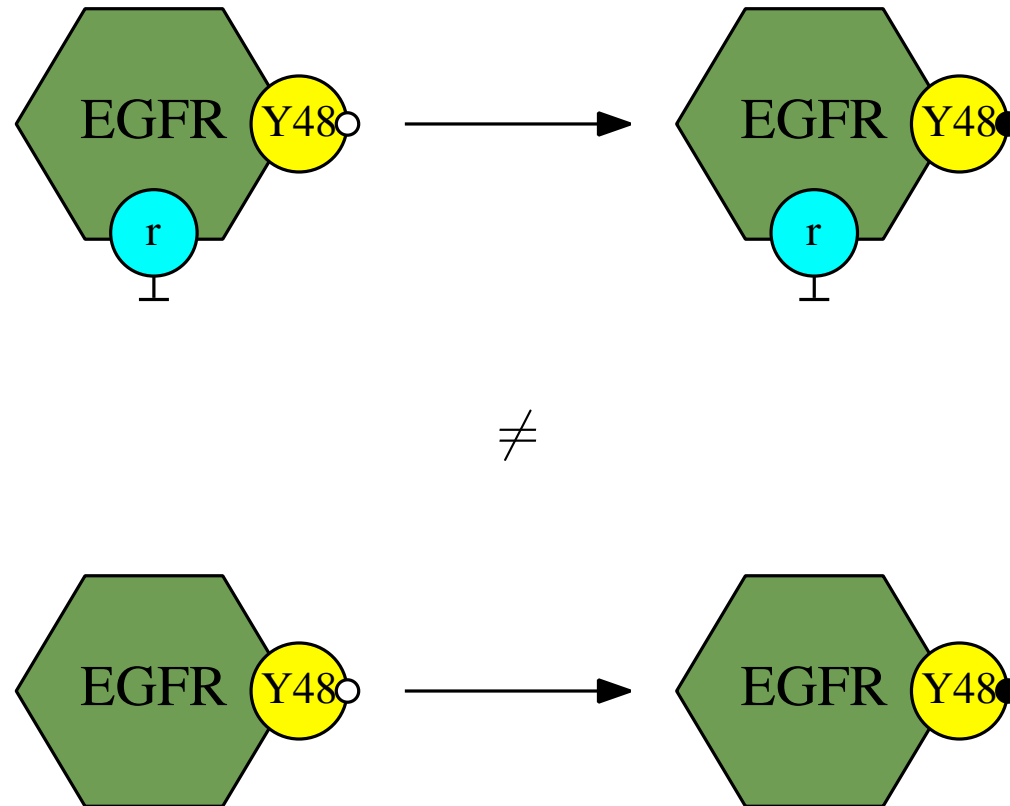
A contextual rule
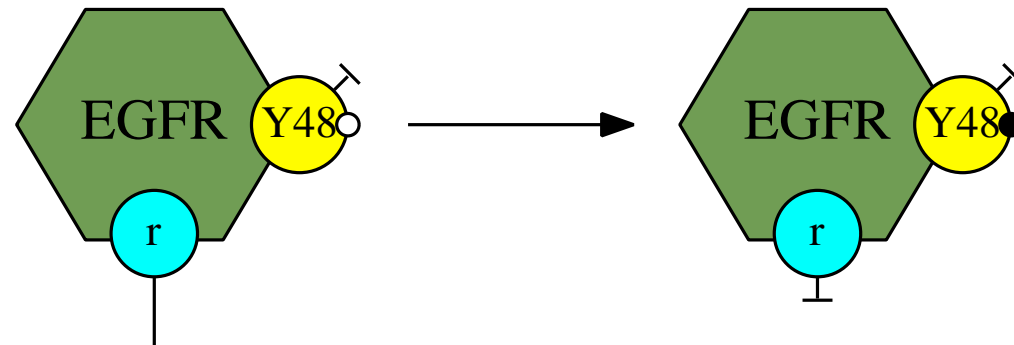
$$EGFR(\text{Y48}{\sim}u,\text{r!\_}) \rightarrow EGFR(\text{Y48}{\sim}p,\text{r})$$

Creation/Suppression
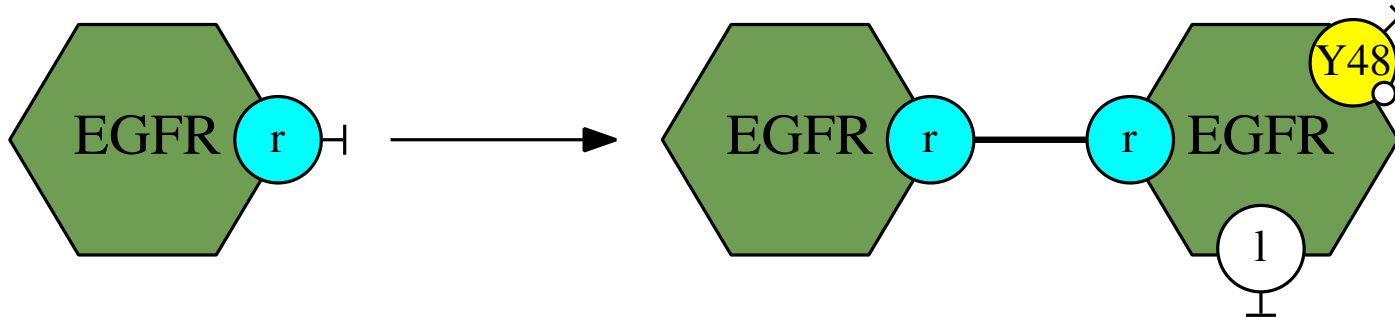
$$R(r) \rightarrow R(r!1), R(r!1,l,Y48{\sim}u)$$
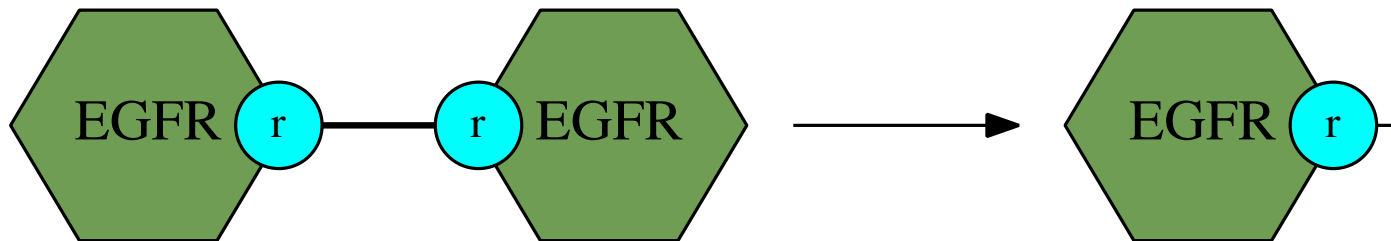
$$R(r!1), R(r!1) \rightarrow R(r)$$

Early EGF example

egf rules 1

protein shorthands: E:=egf, R:=egfr, So:=Sos,Sh:=Sh,G:=grb2
site abbreviations & fusions: Y68:=Y1068, Y48:=Y1148/73, Y7:=Y317, π:=PTB/SH2

- Ligand-receptor binding, receptor dimerisation, rtk x-phosph, & de-phosph
  - 01: $R(l,r), E(r) <-> R(l^1,r), E(r^1)$
  - 02: $R(l^1,r), R(l^2,r) <-> R(l^1,r^3), R(l^2,r^3)$
  - 03: $R(r^1,Y68) -> R(r^1,Y68^p)$
    - $R(Y68^p) -> R(Y68)$
  - 04: $R(r^1,Y48) -> R(r^1,Y48^p)$
    - $R(Y48^p) -> R(Y48)$

receptor type: $R(l,r,Y68,Y48)$

- Sh x-phosph & de-phosph
  - 14: $R(r^2,Y48^{p1}), Sh(\pi^1,Y7) -> R(r^2,Y48^{p1}), Sh(\pi^1,Y7^p)$
  - ??: $Sh(\pi^1,Y7^p) -> Sh(\pi^1,Y7)$
  - 16: $Sh(\pi,Y7^p) -> Sh(\pi,Y7)$

  refined from
  $Sh(Y7^p)-> Sh(Y7)$

- Y68-G binding
  - 09: $R(Y68^p), G(a,b) <-> R(Y68^{p1})+G(a^1,b)$
  - 11: $R(Y68^p), G(a,b^2) <-> R(Y68^{p1})+G(a^1,b^2)$

  refined from
  $R(Y68^p)+G(a)<->R(Y68^{p1})+G(a^1)$

Early EGF example

egf rules 2

refined from
So(d)+G(b)<->So(d^1)+G(b^1)

interface note: highlight
the interacting parts

- **G-So binding**
  - 10: R(Y68^{p1}), G(a^1,b), So(d) <-> R(Y68^{p1}), G(a^1,b^2), So(d^2)
  - 12: G(a,b), So(d) <-> G(a,b^1), So(d^1)
  - 22: Sh(π,Y7^{p2}), G(a^2,b), So(d) <-> Sh(π,Y7^{p2}), G(a^2,b^1), S(d^1)
  - 19: Sh(π^1,Y7^{p2}), G(a^2,b), So(d) <-> Sh(π^1,Y7^{p2}), G(a^2,b^1), S(d^1)

- **Y48-Sh binding**
  - 13: R(Y48^p), Sh(π,Y7) <-> R(Y48^{p1}), Sh(π^1,Y7)
  - 15: R(Y48^p), Sh(π,Y7^p) <-> R(Y48^{p1}), Sh(π^1,Y7^p)
  - 18: R(Y48^p), Sh(π,Y7^{p1}), G(a^1,b) <-> R(Y48^{p2}), Sh(π^2,Y7^{p1}), G(a^1,b)
  - 20: R(Y48^p), Sh(π,Y7^{p1}), G(a^1,b^3), S(d^3) <-> R(Y48^{p2}), Sh(π^2,Y7^{p1}), G(a^1,b^3), S(d^3)

refined from
R(Y48^p)+Sh(π)<->R(Y48^{p1})+Sh(π^1)

why not simply G(b^3)??

- **Sh-G binding**
  - 17: R(Y48^{p1}), Sh(π^1,Y7^p), G(a,b) <-> R(Y48^{p1}), Sh(π^1,Y7^{p2}), G(a^2,b)
  - 21: Sh(π,Y7^p), G(a,b) <-> Sh(π,Y7^{p1}), G(a^1,b)
  - 23: Sh(π,Y7^p), G(a,b^2) <-> Sh(π,Y7^{p1}), G(a^1,b^2)
  - 24: R(Y48^{p1}), Sh(π^1,Y7^p), G(a,b^3), S(d^3) <-> R(Y48^{p1}), Sh(π^1,Y7^{p2}), G(a^2, b^3), S(d^3)

refined from
Sh(π), G(a)<->Sh(π^1), G(a^1)

Properties of interest

1. Show the absence of modeling errors:

   - detect dead rules ;
   - detect overlapping rules;
   - detect non exhaustive interactions;
   - detect rules with ambiguous molecularity.

2. Get idiomatic description of the networks:

   - capture causality;
   - capture potential interactions;
   - capture relationships between site states;
   - simplify rules.

3. Allow fast simulation:

   - capture accurate approximation of the wake-up relation.

Overview

1. Introduction

2. Language: Kappa

3. Abstraction: Local views

4. Completeness: false positives?

5. Local fragment of Kappa

6. Decontextualization

7. Conclusion

Concrete semantics

A rule is a symbolic representation of a multi-set of reactions.

For instance, the rule:
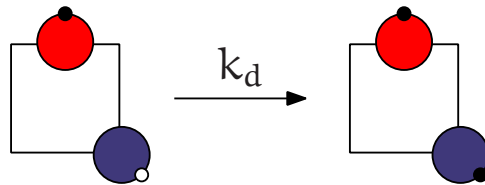
within a model with the following signature:
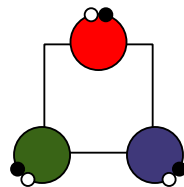
denotes the following two rules:
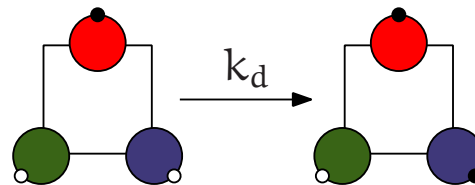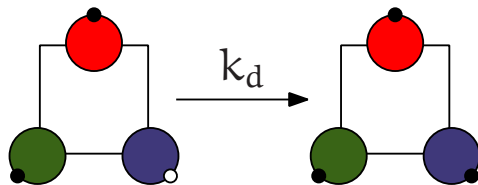
Set of reachable chemical species

Let $\mathcal{R} = \{R_i\}$ be a set of rules.
Let *Species* be the set of all chemical species ($C, c_1, c_1', \dots, c_k, c_k', \dots \in$ *Species*).
Let *Species*$_0$ be the set of initial .

We are interested in *Species*$_\omega$ the set of all chemical species that can be constructed in one or several applications of the reactions induced by the rules in $\mathcal{R}$, starting from the set *Species*$_0$ of initial chemical species.

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

Inductive definition

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

$$
\mathbb{F} : \begin{cases} \wp(\textit{Species}) & \rightarrow \wp(\textit{Species}) \\ X & \mapsto X \cup \left\{ c_j' \;\middle|\; \begin{array}{l} \exists R_k \in \mathcal{R}, c_1, \ldots, c_m \in X, \\ c_1, \ldots, c_m \rightarrow_{R_k} c_1', \ldots, c_n' \end{array} \right\} \end{cases} .
$$

We define the set of reachable chemical species as follows:

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

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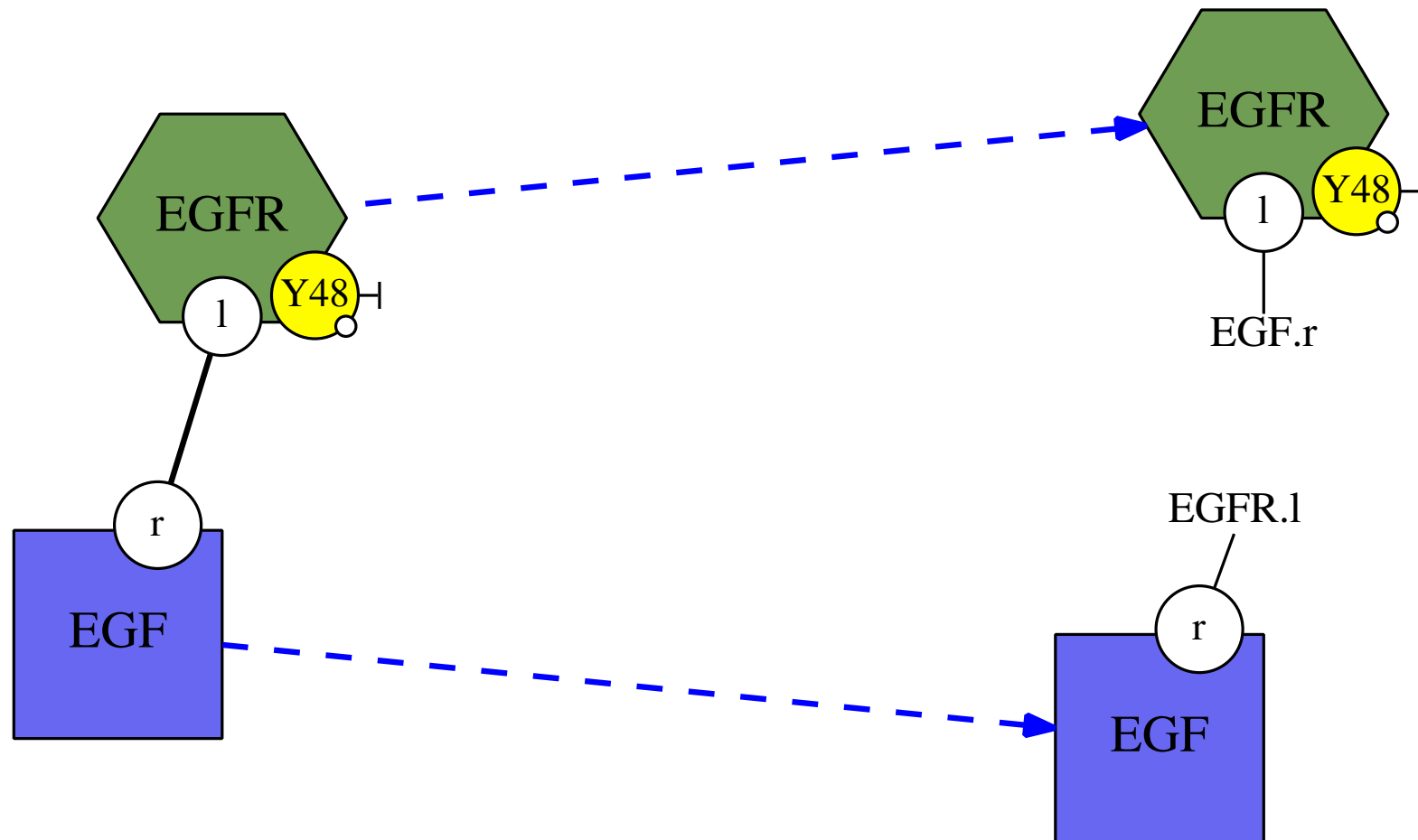
$$\alpha(\{R(Y1\sim u,l!1),\ E(r!1)\}) = \{R(Y1\sim u,l!r.E);\ E(r!l.R)\}.$$

Galois connection

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

- Let $\alpha \in \wp(Species) \to \wp(Local\_view)$ be the function that maps any set of chemical species into the set of their local views.

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

- Let $\gamma \in \wp(Local\_view) \to \wp(Species)$ be the function that maps any set of local views into the set of chemical species that can be built with these local views.

- The pair (α, γ) forms a Galois connection:

$$\wp(Species) \xleftarrow[\alpha]{\gamma} \wp(Local\_view).$$

$\gamma \circ \alpha$

$\gamma \circ \alpha$ is an upper closure operator: it abstracts away some information.

Guess the image of the following set of chemical species ?
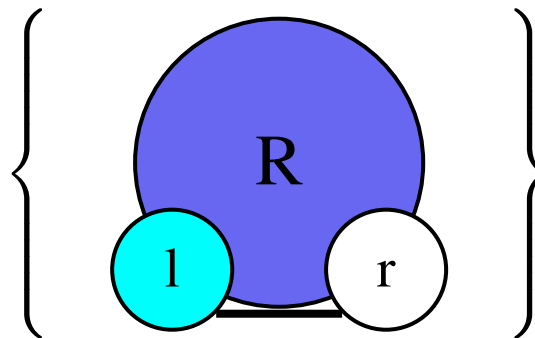
$\alpha \circ \gamma$

$\alpha \circ \gamma$ is a lower closure operator: it simplifies (or reduces) constraints.

Guess the image of the following set of local views ?
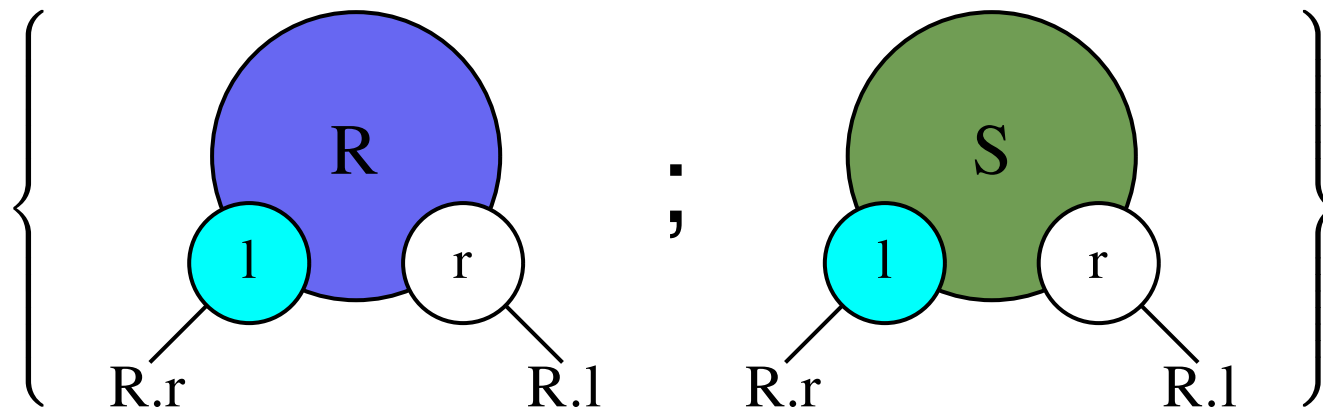
One more question

$\alpha \circ \gamma$ is a lower closure operator: it simplifies (or reduces) constraints.

Guess the image of the following set of local views ?

Jérôme Feret
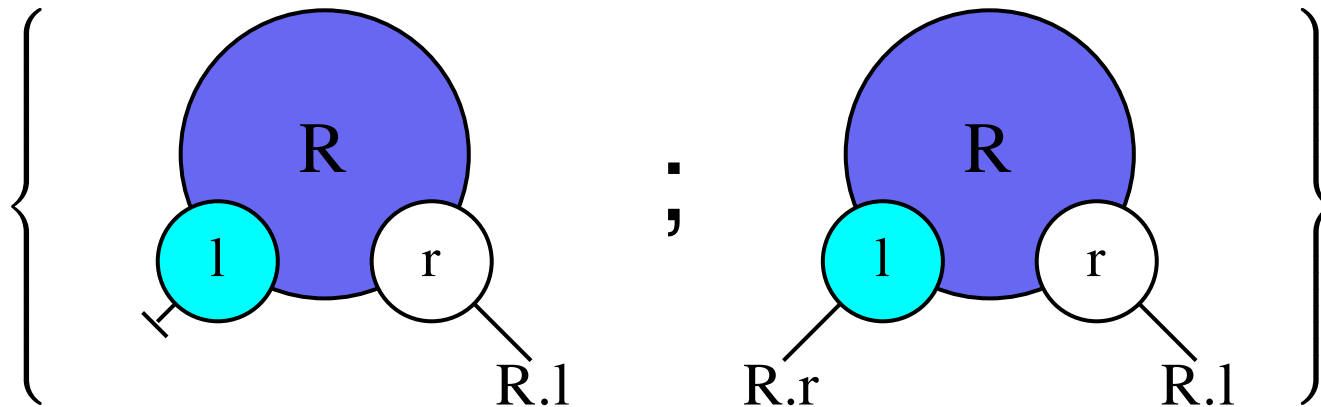
Abstract reactions
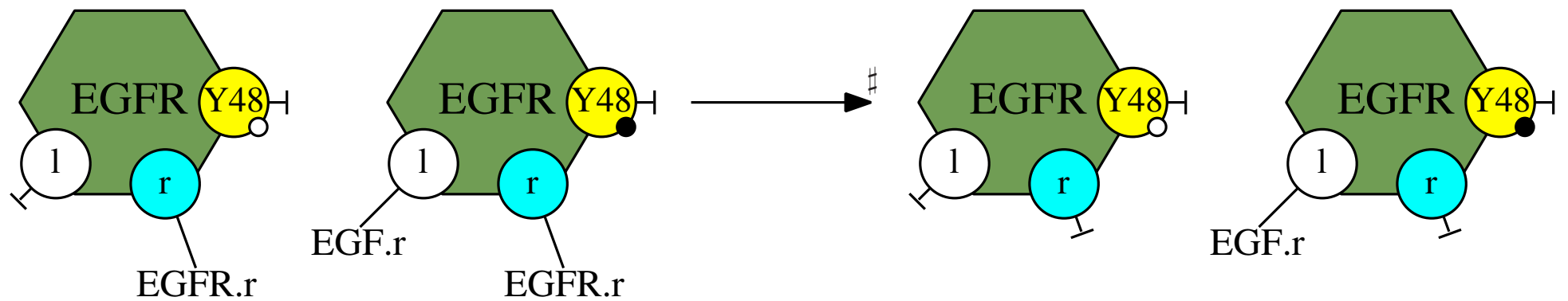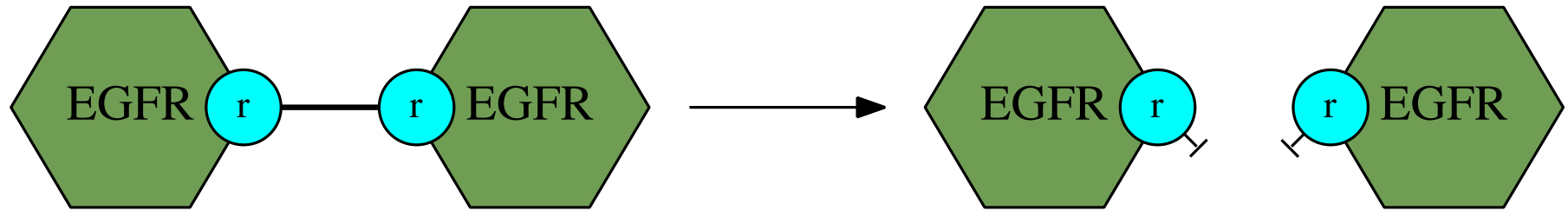
Abstract counterpart to $\mathbb{F}$

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

$$\mathbb{F}^\sharp : \begin{cases} \wp(\mathit{Local\_view}) & \to \wp(\mathit{Local\_view}) \\ Y & \mapsto Y \cup \left\{ lv'_j \;\middle|\; \begin{array}{l} \exists R_k \in \mathcal{R}, lv_1, \ldots, lv_m \in Y, \\ lv_1, \ldots, lv_m \to^\sharp_{R_k} lv'_1, \ldots, lv'_n \end{array} \right\} \end{cases} .$$

Theorem 1 (soundness) It follows that:

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. Language: Kappa

3. Abstraction: Local views

4. Completeness: false positives?

5. Local fragment of Kappa

6. Decontextualization

7. Conclusion

From views to species

For any $X \in \wp(\textit{Local\_view})$, $\gamma(X)$ is given by a rewrite system:
For any $\textit{lv} \in X$, we add the following rules:
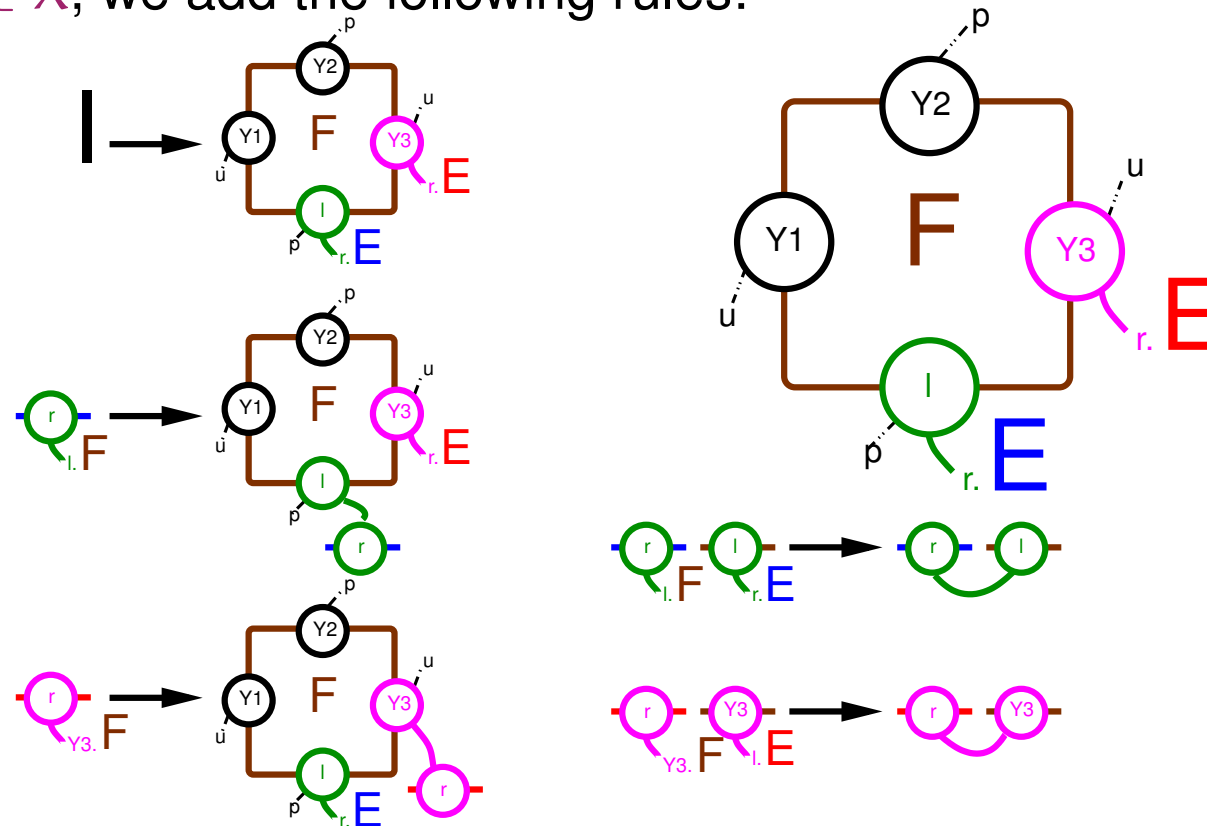
I and semi-links are non-terminal.
I is the initial symbol.

Pumping lemma

- We use this rewrite system to enumerate the chemical species of $\gamma(X)$.

- There are two cases:

  1. either there is a finite number of rewrite sequences;
  2. or we encounter cyclic derivations
     i.e. an open chemical species with a cycle of the following form:
     $$R.l\text{-}r.E \ \ldots \ R.l\text{-}r.E$$

     can be built.

- We only enumerate chemical species that are reached through an acyclic rewriting computation.

- It turns out that: if $X \in \alpha(\wp(\textit{Species}))$ then each rewrite sequence is the prefix of a terminating rewrite sequence.
  (So there is an unbounded number of species if, and only if, there is an unbounded number of rewrite sequences.)

Examples

1. Make the demo for egf
2. Make the demo for fgf
3. Make the demo for Global invariants

Which information is abstracted away ?

Our analysis is exact (no false positive):

- for EGF cascade (356 chemical species);

- for FGF cascade (79080 chemical species);

- for SBF cascade (around 10^{19} chemical species).

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 ?

Which information is abstracted away ?

Theorem 2 We suppose that:

1. $(D, \subseteq)$ be a partial order;

2. $(D^\sharp, \sqsubseteq, \sqcup)$ be chain-complete partial order;

3. $D \xleftarrow[\alpha]{\gamma} D^\sharp$ be a Galois connection;

4. $\mathbb{F} \in D \to D$ and $\mathbb{F}^\sharp \in D^\sharp \to D^\sharp$ are monotonic;

5. $\mathbb{F} \circ \gamma \mathrel{\dot{\subseteq}} \gamma \circ \mathbb{F}^\sharp$;

6. $X_0, inv \in D$ such that:

   - $X_0 \subseteq \mathbb{F}(X_0) \subseteq \mathbb{F}(inv) \subseteq inv$,
   - $inv = \gamma(\alpha(inv))$,
   - and $\alpha(\mathbb{F}(inv)) = \mathbb{F}^\sharp(\alpha(inv))$;

Then, $\mathit{lfp}_{\alpha(X_0)} \mathbb{F}^\sharp$ exists and $\gamma(\mathit{lfp}_{\alpha(X_0)} \mathbb{F}^\sharp) \subseteq inv$.

Species

inv

$\gamma(\mathit{lfp}_{\alpha(Species_0)} \mathbb{F}^\sharp)$

$Species_\omega$
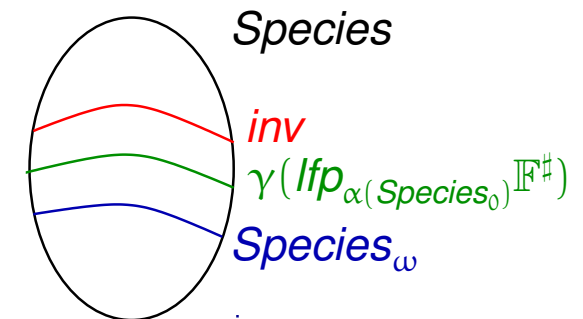
When is there no false positive ?

Theorem 3 We suppose that:

1. $(D, \subseteq, \cup)$ and $(D^\sharp, \sqsubseteq, \sqcup)$ are chain-complete partial orders;

2. $(D, \subseteq) \xleftarrow[\alpha]{\gamma} (D^\sharp, \sqsubseteq)$ is a Galois connection;

3. $\mathbb{F} : D \to D$ is a monotonic map;

4. X_0 is a concrete element such that $X_0 \subseteq \mathbb{F}(X_0)$;

5. $\mathbb{F} \circ \gamma \mathrel{\dot{\subseteq}} \gamma \circ \mathbb{F}^\sharp$;

6. $\mathbb{F}^\sharp \circ \alpha = \alpha \circ \mathbb{F} \circ \gamma \circ \alpha$.

Then:

- $\mathit{lfp}_{X_0} \mathbb{F}$ and $\mathit{lfp}_{\alpha(X_0)} \mathbb{F}^\sharp$ exist;
- $\mathit{lfp}_{X_0} \mathbb{F} = \gamma(\alpha(\mathit{lfp}_{X_0} \mathbb{F})) \iff \mathit{lfp}_{X_0} \mathbb{F} = \gamma(\mathit{lfp}_{\alpha(X_0)} \mathbb{F}^\sharp)$.

We need to understand under which assumptions $\mathit{lfp}_{X_0} \mathbb{F} = \gamma(\alpha(\mathit{lfp}_{X_0} \mathbb{F}))$.

Local set of chemical species

Definition 1 We say that a set $X \in \wp(\textit{Species})$ of chemical species is local if and only if $X \in \gamma(\wp(\textit{Local\_view}))$.

(ie. a set X is local if and only if X is exactly the set of all the species that are generated by a given set of local views.)

Swapping relation

We define the binary relation $\overset{\text{SWAP}}{\sim}$ among tuples *Species** of chemical species. We say that $(C_1, \ldots, C_m) \overset{\text{SWAP}}{\sim} (D_1, \ldots, D_n)$ if and only if:

$(C_1, \ldots, C_m)$ matches with
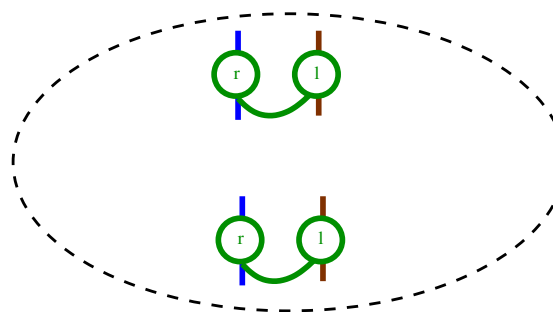
while $(D_1, \ldots, D_n)$ matches with
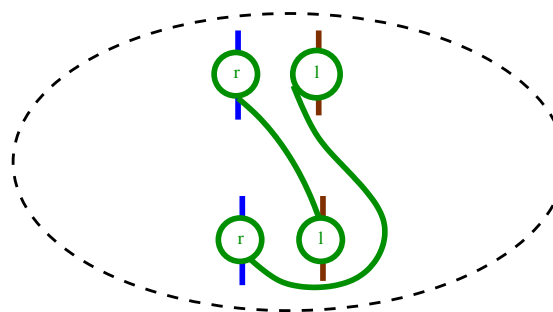
Swapping closure

Theorem 4 Let $X \in \wp(\textit{Species})$ be a set of chemical species.

The two following assertions are equivalent:

1. $X = \gamma(\alpha(X))$;

2. for any tuples $(C_i), (D_j) \in \textit{Species}^*$ such that:

   - $(C_i) \in X^*$,

   - and $(C_i) \overset{\text{SWAP}}{\sim} (D_j)$;

   we have $(D_j) \in X^*$.

Proof (easier implication way)

If:

- $X = \gamma(\alpha(X))$,

- $(C_i)_{i \in I} \in X^*$,

- and $(C_i)_{i \in I} \overset{\text{SWAP}}{\sim} (D_j)_{j \in J}$;

Then:

we have $\alpha(\{C_i \mid i \in I\}) = \alpha(\{D_j \mid j \in J\})$ (because $(C_i) \overset{\text{SWAP}}{\sim} (D_j)$)

and $\alpha(\{C_i \mid i \in I\}) \subseteq \alpha(X)$ (because $(C_i) \in X^*$ and α mon);

so $\alpha(\{D_j \mid j \in J\}) \subseteq \alpha(X)$;

so $\{D_j \mid j \in J\} \subseteq \gamma(\alpha(X))$ (by def. of Galois connections);

so $\{D_j \mid j \in J\} \subseteq X$ (since $X = \gamma(\alpha(X))$);

so $(D_j)_{j \in J} \in X^*$.

Proof: more difficult implication way

For any $X \in \wp(\textit{Local\_view})$, $\gamma(X)$ is given by a rewrite system:
For any $\textit{lv} \in X$, we add the following rules:
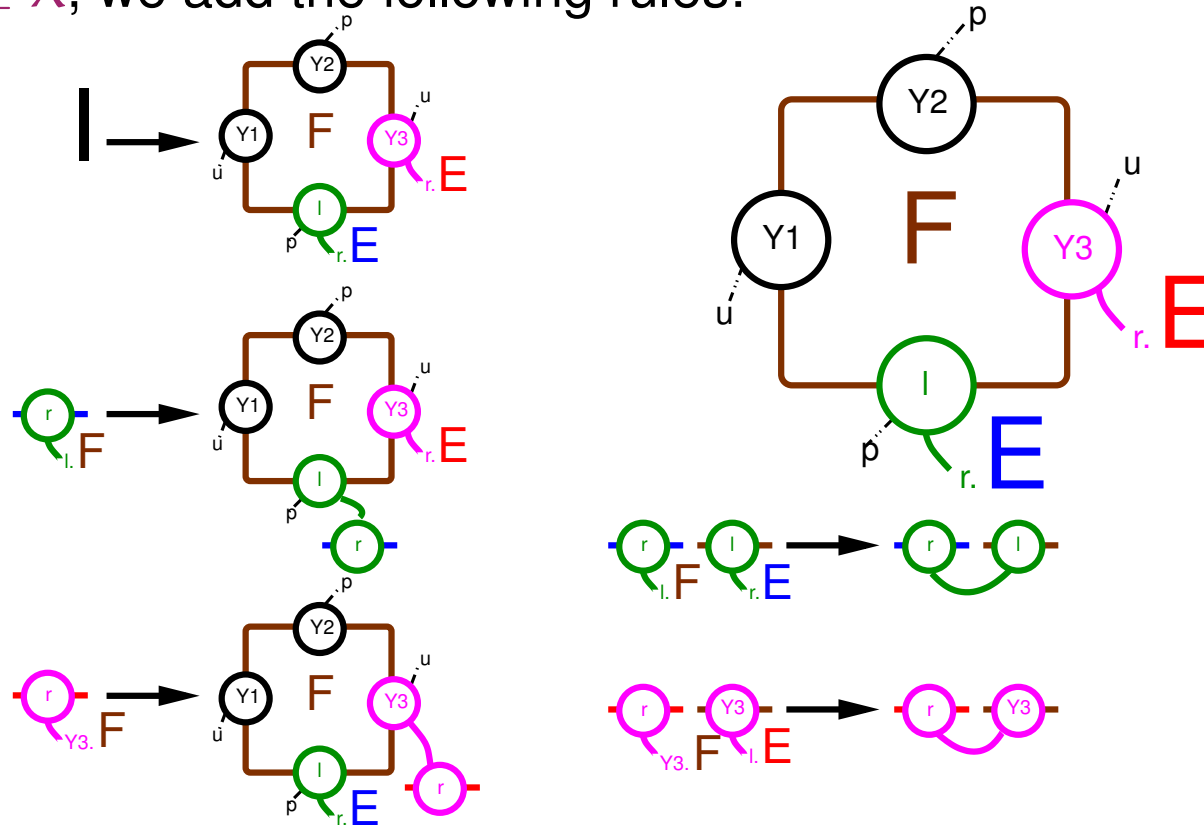
I and semi-links are non-terminal.
I is the initial symbol.

Proof (more difficult implication way)

We suppose that X is close with respect to $\overset{\text{SWAP}}{\sim}$.
We want to prove that $\gamma(\alpha(X)) \subseteq X$.

We prove, by induction, that any open complex that can be built by gathering the views of $\alpha(X)$, can be embedded in a complex in X:

- By def. of α, this is satisfied for any local view in $\alpha(X)$;

- This remains satisfied after unfolding a semi-link with a local view;

- This remains satisfied after binding two semi-links.

Overview

Outline

We have proved that:

- if the set $Species_\omega$ of reachable chemical species is close with respect swapping $\overset{\text{SWAP}}{\sim}$,

- then the reachability analysis is exact (i.e. $Species_\omega = \gamma(lfp_{\alpha(Species_0)} \mathbb{F}^\sharp)$).

Now we give some sufficient conditions that ensure this property.

Sufficient conditions

Whenever the following assumptions:

1. initial agents are not bound;

2. rules are atomic;

3. rules are local:

   - only agents that interact are tested,
   - no cyclic patterns (neither in lhs, nor in rhs);

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. chemical species in $\gamma(\alpha(Species_\omega))$ are acyclic,

are satisfied, the set of reachable chemical species is local.

Non local systems

$Species_0 \triangleq$ R(a~u)

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

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

Non local systems

$$Species_0 \triangleq \text{A}(a{\sim}u), \text{B}(a{\sim}u)$$

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

$\text{A}(a{\sim}u!1), \text{B}(a{\sim}p!1) \in Species_\omega$
$\text{A}(a{\sim}p!1), \text{B}(a{\sim}u!1) \in Species_\omega$
But $\text{A}(a{\sim}p!1), \text{B}(a{\sim}p!1) \notin Species_\omega$.

Non local systems

$$Species_0 \triangleq A(a{\sim}u)$$

$$Rules \triangleq \left\{ \begin{array}{l} A(a{\sim}u) \leftrightarrow A(a{\sim}p) \\ A(a{\sim}u), A(a{\sim}p) \rightarrow A(a{\sim}u!1), A(a{\sim}p!1) \end{array} \right\}$$

$A(a{\sim}u!1), A(a{\sim}p!1) \in Species_\omega$
But $A(a{\sim}p!1), A(a{\sim}p!1) \notin Species_\omega$.

Non local systems

$Species_0 \triangleq$ R(a,b)

$Rules \triangleq$ { R(a,b),R(a) $\rightarrow$ R(a,b!1),R(a!1)}

R(a,b!2),R(a!2,b!1),R(a!1,b)$\in Species_\omega$
But R(a!1,b!1) $\notin Species_\omega$.

Overview

1. Introduction

2. Language: Kappa

3. Abstraction: Local views

4. Completeness: false positives?

5. Local fragment of Kappa

6. Decontextualization

7. Conclusion

Outline

- we have a syntactic criterion in order to ensure that the set of reachable chemical species of a kappa system is local ;

- we now design 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 chemical species;
        - uses backtrack to add new rules.

Example

Initial rule:

R2(l!2,r),R1(l!1,r),E2(r!1),E1(r!2) → R2(l!3,r!1),R1(l!2,r!1),E2(r!2),E1(r!3)

Decontextualized rule:

R2(l!_,r),R1(l!_,r) → R2(l!_,r!1),R1(l!_,r!1)

We can remove redundant tests.

Example

Initial rules:

$$Sh(Y7{\sim}p!2,pi!1),G(a!2,b),R(Y48{\sim}p!1) \rightarrow Sh(Y7{\sim}p,pi!1),G(a,b),R(Y48{\sim}p!1)$$
$$Sh(Y7{\sim}p!3,pi!1),G(a!3,b!2),So(d!2),R(Y48{\sim}p!1) \rightarrow Sh(Y7{\sim}p,pi!1),G(a,b!2),So(d!2),R(Y48{\sim}p!1)$$
$$Sh(Y7{\sim}p!1,pi),G(a!1,b) \rightarrow Sh(Y7{\sim}p,pi),G(a,b)$$
$$Sh(Y7{\sim}p!1,pi),G(a!1,b!\_) \rightarrow Sh(Y7{\sim}p,pi),G(a,b!\_)$$

Decontextualized rule:

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

We can remove exhaustive enumerations.

An undecontextualizable rule

Initial rule:

Sh(Y7~u,pi!1),R(Y48~p!1,r!_) -> Sh(Y7~p,pi!1),R(Y48~p!1,r!_)

Decontextualized rule:

Sh(Y7~u,pi!1),R(Y48!1,r!_) -> Sh(Y7~p,pi!1),R(Y48!1,r!_)

Conjugation

If a rule R' is equivalent to a rule in the transitive closure of the system.
Then it may be included in the system without modifying reachable states.
To remove the context C of a rule, we try to apply it for another context C' by:

1. removing the context C' (backtrack) ;

2. building the context C ;

3. applying the initial rule ;

4. removing the context C (backtrack) ;

5. building the context C'.

This is proved manually.

Overview

1. Introduction

2. Language: Kappa

3. Abstraction: Local views

4. Completeness: false positives?

5. Local fragment of Kappa

6. Decontextualization

7. Conclusion

Conclusion

- A scalable static analysis to abstract the reachable chemical species.

- A class of models for which the abstraction is complete.

- Many applications:
  - idiomatic description of reachable chemical species;
  - 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 chemical species);
  - on the fly flat rule generation (for large/unbounded set)