MPRI

Abstract interpretation of protein-protein interactions networks

Jérôme Feret
Laboratoire d’Informatique de l’École Normale Supérieure
INRIA, ÉNS, CNRS
http://www.di.ens.fr/~feret

Wednesday, the 11th of February, 2015
Joint-work with...

Walter Fontana
Harvard Medical School

Vincent Danos
ÉNS

Ferdinanda Camporesi
Bologna / ÉNS

Russ Harmer
ÉNS Lyon

Jean Krivine
Paris VII
Signalling Pathways

EGF, TGF-alpha, etc

EGFR

PI3-K

AKT

phosphorylation

GRB2

SOS

RAS

RAF

ERK

MEK

Gene transcription
Cell cycle progression

Cell proliferation
Inhibition of apoptosis
Angiogenesis
Migration, Adhesion, Invasion

Eikuch, 2007
Bridging the gap between... knowledge representation and models of dynamical systems
Rule-based approach

We use site graph rewrite systems

1. The description level matches with both
   - the observation level
   - and the intervention level
   of the biologist.
   We can tune the model easily.

2. Model description is very compact.

3. Quantitative semantics can be defined.
Complexity walls

The diagram illustrates the relationship between the number of instances per molecular species and the number of molecular species. It categorizes models into:

- Deterministic differential equations
- Stochastic master equations
- Agent/rule-based

The diagram also highlights different system types:

- Early EGF
- EGF to ERK
- EGF to ERK and AKT
- EGF with receptor network to ERK and AKT

The diagram shows a combinatorial wall and an event wall, indicating the complexity barriers in modeling biological systems.
Abstractions offer different perspectives

Concrete semantics

Flow of information

Causal traces modulo abstraction

Exact projection of the ODE semantics
Static analysis of reachable species (I/II)

Semi-fluid medium: the notion of individual is meaningless.

Design a static analysis to approximate the set of reachable species [VMCAI’08] which focuses on the relationships between the states of the sites of each agent:

This analysis is efficient, suitable to our problem, and accurate.
Static analysis of reachable species (II/II)

Applications:

1. check the consistency of a model [ICCMSE’07]
2. compute the properties to allow fast simulation [APLAS’07]
3. simplify models,
4. compute independent fragments of chemical species [PNAS’09, LICS’10, Chaos’10]

The analysis is complete (no false positif) for a significatif kernel of Kappa [VMCAI’08].
Model reduction

The ground differential system uses one variable per chemical species; We directly compute its exact projection over independent fragments of chemical species.

With a small model, 356 chemical species are reduced into 38 fragments:

On a bigger model, $10^{19}$ chemical species are reduced into $180\,000$ fragments. [PNAS’09, LICS’10, Chaos’10]
Reachability Analysis of Rule-based Models

[ICCMSE'07, VMCAI'08]

Jérôme Feret
Département d’Informatique de l’École Normale Supérieure
INRIA, ÉNS, CNRS
http://www.di.ens.fr/~feret

Wednesday, the 11th of February, 2015
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
A single story
A concurrent story
Overshoot

When we combine the two stories...
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

\[ E(r!1), R(l!1, r!2), R(r!2, l!3), E(r!3) \]
A Unbinding/Binding Rule

\[ E(r), \ R(l,r) \leftrightarrow E(r!1), \ R(l!1,r) \]
Internal state

\[ R(Y_1 \sim u, l!1), \ E(r!1) \leftrightarrow R(Y_1 \sim p, l!1), \ E(r!1) \]
Early EGF example

**egf rules 1**

- **Ligand-receptor binding, receptor dimerisation, rtk x-phosph, & de-phosph**
  - 01: $R(l,r), E(r) \leftrightarrow R(l',r), E(r')$
  - 02: $R(l',r), R(l'',r) \leftrightarrow R(l',r''), R(l'',r'')$
  - 03: $R(r^1,Y68) \rightarrow R(r^1,Y68^p)$
    - $R(Y68^p) \rightarrow R(Y68)$
  - 04: $R(r^1,Y48) \rightarrow R(r^1,Y48^p)$
    - $R(Y48^p) \rightarrow R(Y48)$

- **Sh x-phosph & de-phosph**
  - 14: $R(r^2,Y48^{pl}), Sh(\pi, Y7) \rightarrow R(r^2,Y48^{pl}), Sh(\pi, Y7^p)$
  - 15: $Sh(\pi, Y7^p) \rightarrow Sh(\pi, Y7)$
    - refined from $Sh(Y7^p) \rightarrow Sh(Y7)$

- **Y68-G binding**
  - 09: $R(Y68^p), G(a,b) \leftrightarrow R(Y68^{pl}) + G(a^1,b)$
  - 11: $R(Y68^p), G(a,b^2) \leftrightarrow R(Y68^{pl}) + G(a^1,b^2)$
    - refined from $R(Y68^p) + G(a) \leftrightarrow R(Y68^{pl}) + G(a^1)$

**protein shorthands:** E: =egf, R: =egfr, Sos: =Sos, Sh: =Sh, G: =grb2

**site abbreviations & fusions:** Y68:=Y1068, Y48:=Y1148, Y7:=Y317, π:=PTB/SH2

**receptor type:** $R(l,r,Y68,Y48)$
Early EGF example

eGF rules 2

G-So binding
- 10: $R(Y68^{p1}), G(a_1, b), So(d) \leftrightarrow R(Y68^{p1}), G(a_1, b^2), So(d^2)$
- 12: $G(a, b), So(d) \leftrightarrow G(a, b^1), So(d^1)$
- 22: $Sh(\pi, Y7^{p2}), G(a_2, b), So(d) \leftrightarrow Sh(\pi, Y7^{p2}), G(a_2, b^1), S(d^1)$
- 19: $Sh(\pi_1, Y7^{p2}), G(a_2, b), So(d) \leftrightarrow Sh(\pi_1, Y7^{p2}), G(a_2, b^1), S(d^1)$

Y48-Sh binding
- 13: $R(Y48^{p1}), Sh(\pi, Y7) \leftrightarrow R(Y48^{p1}), Sh(\pi, Y7)$
- 15: $R(Y48^{p1}), Sh(\pi, Y7^{p1}) \leftrightarrow R(Y48^{p1}), Sh(\pi, Y7^{p1})$
- 18: $R(Y48^{p1}), Sh(\pi, Y7^{p1}), G(a_1, b) \leftrightarrow R(Y48^{p2}, Sh(\pi, Y7^{p1}), G(a_1, b)$
- 20: $R(Y48^{p1}), Sh(\pi, Y7^{p1}), G(a_1, b^3), S(d^3) \leftrightarrow R(Y48^{p2}, Sh(\pi, Y7^{p1}), G(a_1, b^3), S(d^3)$

Sh-G binding
- 17: $R(Y48^{p1}), Sh(\pi, Y7^{p1}), G(a, b) \leftrightarrow R(Y48^{p1}), Sh(\pi, Y7^{p1}), G(a_2, b)$
- 21: $Sh(\pi, Y7^{p1}), G(a, b) \leftrightarrow Sh(\pi, Y7^{p1}), G(a_1, b)$
- 23: $Sh(\pi, Y7^{p1}), G(a, b^2) \leftrightarrow Sh(\pi, Y7^{p1}), G(a, b^2)$
- 24: $R(Y48^{p1}), Sh(\pi, Y7^{p1}), G(a_1, b^3), S(d^3) \leftrightarrow R(Y48^{p1}), Sh(\pi, Y7^{p1}), G(a_2, b^3), S(d^3)$
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
We write $Z \triangleleft_{\Phi} Z'$ iff:

- $\Phi$ is a site-graph morphism:
  - $i$ is less specific than $\Phi(i)$,
  - if there is a link between $(i, s)$ and $(i', s')$,
    then there is a link between $(\Phi(i), s)$ and $(\Phi(i'), s')$.

- $\Phi$ is an into map (injective):
  - $\Phi(i) = \Phi(i')$ implies that $i = i'$. 

Jérôme Feret 17 Wednesday, the 11th of February, 2015
Set of reachable chemical species

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

$$c_1, \ldots, c_m \rightarrow_{R_k} c_1', \ldots, c_n'$$

whenever:

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

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

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

We define the mapping $F$ as follows:

$$F : \begin{cases} \varphi(Species) & \rightarrow \varphi(Species) \\ X & \mapsto X \cup \left\{ c'_j \mid \exists R_k \in \mathcal{R}, c_1, \ldots, c_m \in X, c_1, \ldots, c_m \rightarrow_{R_k} c'_1, \ldots, c'_n \right\} \end{cases}.$$  

The set $\varphi(Species)$ is a complete lattice.
The mapping $F$ is an extensive $\cup$-complete morphism.

We define the set of reachable chemical species as follows:

$$Species_\omega = \bigcup \{ F^n(Species_0) \mid n \in \mathbb{N} \}.$$
Local views

\[ \alpha(\{R(Y1 \sim u,l!1), E(r!1)\}) = \{R(Y1 \sim u,l!r.E); E(r!!.R)\}. \]
Galois connection

Let $\mathit{Local\_view}$ be the set of all local views.

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

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

Thus, it defines a Galois connection:

$$\wp(\mathit{Species}) \overset{\gamma}{\leftarrow} \overset{\alpha}{\rightarrow} \wp(\mathit{Local\_view}).$$

(The function $\gamma$ maps a set of local views into the set of complexes that can be built with these local views).
$\gamma \circ \alpha$ is an upper closure operator: it abstracts away some information.

Guess the image of the following set of chemical species?
\( \alpha \circ \gamma \) is a lower closure operator: it simplifies (or reduces) constraints.

Guess the image of the following set of local views?

\[ \{ R, S \} \]
$\alpha \circ \gamma$ is a lower closure operator: it simplifies (or reduces) constraints.

Guess the image of the following set of local views?

\[
\begin{align*}
\{ & R \} : \{ & R \}
\end{align*}
\]
Abstract reactions
Abstract counterpart to $F$

We define $F^\#$ as:

$$F^\#: \begin{cases} \varphi(Local\_view) & \rightarrow \varphi(Local\_view) \\ Y & \mapsto Y \cup \left\{ l' \left| \exists R_k \in \mathcal{R}, l'_1, \ldots, l'_{m} \in Y, l'_1, \ldots, l'_m \xrightarrow{\#} R_k \right. \right. \left. \left. l'_1, \ldots, l'_n \right\} \right. \end{cases}$$

We have:

- $F^\#$ is extensive;
- $F^\#$ is monotonic;
- $F \circ \gamma \subseteq \gamma \circ F^\#$;
- $F^\# \circ \alpha = \alpha \circ F \circ \gamma \circ \alpha$ (we will see later why).
Soundness

Theorem 1 Let:

1. \((D, \subseteq, \cup)\) and \((D^\#, \subseteq, \cup)\) be chain-complete partial orders;
2. \(D \xleftarrow{\gamma} D^\#\) be a Galois connection;
3. \(F \in D \rightarrow D\) and \(F^\# \in D^\# \rightarrow D^\#\) be monotonic mappings such that:
   \(F \circ \gamma \subseteq \gamma \circ F^\#\);
4. \(X_0 \in D\) be an element such that: \(X_0 \subseteq F(X_0)\);

Then:

1. both \(lfp_{X_0}F\) and \(lfp_{\alpha(X_0)}F^\#\) exist,
2. \(lfp_{X_0}F \subseteq \gamma(lfp_{\alpha(X_0)}F^\#)\).
Overview

1. Introduction
2. Language: Kappa
3. Abstraction: Local views
4. Completeness: false positives?
5. Local fragment of Kappa
6. Decontextualization
7. Conclusion
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. \((D^\#, \subseteq, \sqcup)\) be chain-complete partial order;
3. \(D \xleftarrow{\gamma} D^\#\) be a Galois connection;
4. \(F \in D \rightarrow D\) and \(F^\# \in D^\# \rightarrow D^\#\) are monotonic;
5. \(F \circ \gamma \subseteq \gamma \circ F^\#\);
6. \(X_0, \text{inv} \in D\) such that:
   - \(X_0 \subseteq F(X_0) \subseteq F(\text{inv}) \subseteq \text{inv}\),
   - \(\text{inv} = \gamma(\alpha(\text{inv}))\),
   - and \(\alpha(F(\text{inv})) = F^#(\alpha(\text{inv}))\);

Then, \(lfp_{\alpha(X_0)F^#}\) exists and \(\gamma(lfp_{\alpha(X_0)F^#}) \subseteq \text{inv}\).
We have already seen (previous lectures) that:

1. $\text{lfp}_{\alpha(X_0)}F^\#$ exists;

2. there exists an ordinal $\delta$ such that $\text{lfp}_{\alpha(X_0)}F^\# = F^\#\delta(\alpha(X_0))$. 

Proof I/III
Proof II/III

Let us show that $\gamma(lfp_{\alpha(X_0)}F^\#) \subseteq inv$.

Let us prove instead by induction over $\delta$ that $F^\#(\alpha(X_0)) \subseteq \alpha(inv)$.

- If $Y \in D^\#$ is an element such that $Y \subseteq \alpha(inv)$,
  $F^\#(Y) \subseteq F^\#(\alpha(inv))$ (F is mon)
  $F^\#(\alpha(inv)) = \alpha(F(inv))$ (assumption)
  $\alpha(F(inv)) \subseteq \alpha(inv)$. (\alpha is mon and inv is a post)

  Thus: $F^\#(Y) \subseteq \alpha(inv)$

- If $Y_i \in D^\#_I$ is a chain of elements such that $Y_i \subseteq \alpha(inv)$ for any $i \in I$,
  then, $\bigsqcup Y_i \subseteq \alpha(inv)$ (lub).

So: $F^\#(\alpha(X_0)) \subseteq \alpha(inv)$. 
Proof III/III

We have:

\[ \mathbb{F}^{\#\delta}(\alpha(X_0)) \subseteq \alpha(inv). \]

Since \( \gamma \) is monotonic:

\[ \gamma(\mathbb{F}^{\#\delta}(\alpha(X_0))) \subseteq \gamma(\alpha(inv)). \]

But, by assumption, \( \gamma(\alpha(inv)) = inv \).

Thus,

\[ \gamma(\mathbb{F}^{\#\delta}(\alpha(X_0))) \subseteq inv. \]
When is there no false positive?

**Theorem 3** We suppose that:

1. $(D, \subseteq, \cup)$ and $(D^\#, \subseteq, \cup)$ are chain-complete partial orders;
2. $(D, \subseteq) \xrightarrow{\gamma} (D^\#, \subseteq)$ is a Galois connection;
3. $\mathcal{F} : D \rightarrow D$ is a monotonic map;
4. $X_0$ is a concrete element such that $X_0 \subseteq \mathcal{F}(X_0)$;
5. $\mathcal{F} \circ \gamma \subseteq \gamma \circ \mathcal{F}^#$;
6. $\mathcal{F}^# \circ \alpha = \alpha \circ \mathcal{F} \circ \gamma \circ \alpha$.

Then:

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

We need to understand under which assumptions $\text{lfp}_{X_0}F = \gamma(\alpha(\text{lfp}_{X_0}F))$. 
We define the binary relation \( \sim_{\text{SWAP}} \) among tuples \( \text{Species}^* \) of chemical species. We say that \((C_1, \ldots, C_m) \sim_{\text{SWAP}} (D_1, \ldots, D_n)\) if and only if:

\[
(C_1, \ldots, C_m) \text{ matches with } (C_1, \ldots, C_m)
\]

while \((D_1, \ldots, D_n)\) matches with

\[
(D_1, \ldots, D_n)
\]
Swapping closure

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

The two following assertions are equivalent:

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

2. for any tuples $(C_i), (D_j) \in Species^*$ 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 $\alpha$ 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 $X \in \mathcal{P}(Local\_view)$, $\gamma(X)$ is given by a rewrite system:
For any $lv \in X$, 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 $\sim^{\text{SWAP}}$. 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 $\alpha$, 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.
Initialization

\[ \forall \ell \in \alpha(X) \]
Unfolding a semi-link

open partial species

C ∈ X

C′ ∈ X
Unfolding a semi-link

open partial species

C'' ∈ X

SWAP

~

Jérôme Feret

Wednesday, the 11th of February, 2015
Binding two semi-links

\[ C \in X \]

\[ C'' \in X \] (SWAP)}
Consequences

Let $Y \in \wp(\text{Local\_view})$ be a set of local views such that $\alpha(\gamma(Y)) = Y$.

1. Each open complex $C$ built with the local views in $Y$ is a sub-complex of a close complex $C'$ in $\gamma(Y)$.

2. When considering the rewrite system that computes $\gamma(Y)$, any partial rewriting sequence can be completed in a successful one.

Thus:

(a) $\gamma(Y)$ is finite if and only if the grammar has a finite set of prefixes (and the latter is decidable);

(b) We have $\mathbb{P}^\sharp \circ \alpha = \alpha \circ \mathbb{F} \circ \gamma \circ \alpha$. 
Overview

1. Introduction
2. Language: Kappa
3. Abstraction: Local views
4. Completeness: false positives?
5. Local fragment of Kappa
6. Decontextualization
7. Conclusion
We have proved that:

- if the set $\text{Species}_\omega$ of reachable chemical species is close with respect to swapping $\sim$,
- then the reachability analysis is exact (i.e. $\text{Species}_\omega = \gamma(\text{lfp}_\alpha(\text{Species}_0)^{F^g})$).

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:
   - $A(a \sim m, S), B(b \sim n, T) \rightarrow A(a \sim m!1, S), B(b \sim n!1, T)$
   - and $A(a \sim m', S'), B(b \sim n', T') \rightarrow A(a \sim m'!1, S'), B(b \sim n'!1, T')$,
   then:
   - $A(a \sim m, S), B(b \sim n', T') \rightarrow A(a \sim m!1, S), B(b \sim n'!1, T')$;
5. chemical species in $\gamma(\alpha(Species_\omega))$ are acyclic,
   are satisfied, the set of reachable chemical species is local.
Proof outline

We sketch a proof in order to discover sufficient conditions that ensure this property:

- We consider tuples of complexes in which the same kind of links occur twice.
- We want to swap these links.
- We introduce the history of their computation.
- There are several cases...
First case (I/V)

\[ C \in \mathit{Species}_\omega \]

\[ C' \in \mathit{Species}_\omega \]
First case (II/V)

just before the links are made

\[ C \in \text{Species}_\omega \]

\[ C' \in \text{Species}^*_\omega \]
First case (III/V)

we suppose we can swap the links

\[ C \in \text{Species}_{\omega}^* \]
First case (IV/V)

Then, we ensure that further computation steps:

- are always possible;
- have the same effect on local views;
- commute with the swapping relation $\sim$.

$$\begin{align*}
\text{C}_n & \xrightarrow{\text{SWAP} \sim, \sigma} \text{C}'_n \\
\text{R}, \phi & \downarrow \quad \text{R}, \phi \\
\text{C}_{n+1} & \xrightarrow{\text{SWAP} \sim, \sigma} \text{C}'_{n+1}
\end{align*}$$
First case (V/V)

C ∈ Species_ω^∗
we assume that the chemical species $C$ is acyclic
Second case (II/II)
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:
   - \( A(a \sim m,S),B(b \sim n,T) \rightarrow A(a \sim m!1,S),B(b \sim n!1,T) \)
   - and \( A(a \sim m',S'),B(b \sim n',T') \rightarrow A(a \sim m'!1,S'),B(b \sim n'!1,T') \),
   then:
     - \( A(a \sim m,S),B(b \sim n',T') \rightarrow A(a \sim m!1,S),B(b \sim n'!1,T') \);
5. chemical species in \( \gamma(\alpha(Species,\omega)) \) are acyclic,
   are satisfied, the set of reachable chemical species is local.
Third case (I/III)

\[ C \in \text{Species}_\omega \]
Third case (II/III)

\[ C \in \text{Species}_\omega^* \]

\[ r \quad r \]
Third case (II/III)

\[ C \in \text{Species}_\omega^* \]
Non local systems

\[\text{Species}_0 \triangleq R(a\sim u)\]

\[\text{Rules} \triangleq \begin{cases} 
R(a\sim u) & \leftrightarrow R(a\sim p) \\
R(a\sim u), R(a\sim u) & \rightarrow R(a\sim u!1), R(a\sim u!1) \\
R(a\sim p), R(a\sim u) & \rightarrow R(a\sim p!1), R(a\sim p!1) \\
R(a\sim p), R(a\sim p) & \rightarrow R(a\sim p!1), R(a\sim p!1) 
\end{cases}\]

\[R(a\sim u!1), R(a\sim u!1) \in \text{Species}_\omega\]

\[R(a\sim p!1), R(a\sim p!1) \in \text{Species}_\omega\]

But \(R(a\sim u!1), R(a\sim p!1) \notin \text{Species}_\omega\).
Non local systems

\[ \text{Species}_0 \triangleq A(\sim u), B(\sim u) \]

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

\[ A(\sim u!1), B(\sim p!1) \in \text{Species}_\omega \]

\[ A(\sim p!1), B(\sim u!1) \in \text{Species}_\omega \]

But \[ A(\sim p!1), B(\sim p!1) \notin \text{Species}_\omega. \]
Non local systems

\[ \text{Species}_0 \triangleq A(a \sim u) \]

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

\[ A(a \sim u!1), A(a \sim p!1) \in \text{Species}_\omega \]

But \[ A(a \sim p!1), A(a \sim p!1) \notin \text{Species}_\omega. \]
Non local systems

\[ \text{Species}_0 \triangleq R(a,b) \]
\[ \text{Rules} \triangleq \{ R(a,b), R(a) \rightarrow R(a,b!1), R(a!1) \} \]

\[ R(a,b!2), R(a!2,b!1), R(a!1,b) \in \text{Species}_\omega \]
But \[ R(a!1,b!1) \notin \text{Species}_\omega. \]
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:

\[ R2(l!2,r), R1(l!1,r), E2(r!1), E1(r!2) \rightarrow R2(l!3,r!1), R1(l!2,r!1), E2(r!2), E1(r!3) \]

Decontextualized rule:

\[ R2(l!_,r), R1(l!_,r) \rightarrow R2(l!_,r!1), R1(l!_,r!1) \]

We can remove redundant tests.
Example

Initial rules:

\[ \text{Sh}(Y_7 \sim p!2, pi!1), G(a!2,b), R(Y_{48} \sim p!1) \rightarrow \text{Sh}(Y_7 \sim p, pi!1), G(a,b), R(Y_{48} \sim p!1) \]
\[ \text{Sh}(Y_7 \sim p!3, pi!1), G(a!3,b!2), \text{So}(d!2), R(Y_{48} \sim p!1) \rightarrow \text{Sh}(Y_7 \sim p, pi!1), G(a,b!2), \text{So}(d!2), R(Y_{48} \sim p!1) \]
\[ \text{Sh}(Y_7 \sim p!1, pi), G(a!1,b) \rightarrow \text{Sh}(Y_7 \sim p, pi), G(a,b) \]
\[ \text{Sh}(Y_7 \sim p!1, pi), G(a!1,b\_}) \rightarrow \text{Sh}(Y_7 \sim p, pi), G(a,b\_) \]

Decontextualized rule:

\[ \text{Sh}(Y_7!1), G(a!1) \rightarrow \text{Sh}(Y_7), G(a) \]

We can remove exhaustive enumerations.
How does it work?

To remove a test, we prove that:

- this test is satisfied whenever the other tests are satisfied;
- or each complex that passes all tests but this one also matches with the left hand side of another rule that performs the same action.
More formally:

- Each rule $R$ is associated with the set $S(R)$ of open chemical species that can match its lhs;
- Rules are gathered in equivalence classes according to the actions they perform;
- For each class $[R]$, we compute:
  \[
  G([R]) = \bigcup \{S(R') | R' \in [R]\}.
  \]
- For each class $[R]$, $Reach([R])$ is an over approximation of the set of open chemical species that may match the lhs of a rule $R' \in [R]$.

A rule $R$ may be decontextualized in a rule $R'$ if:

\[
S(R') \cap Reach([R]) \subseteq G([R]).
\]

Decontextualization is more efficient, if the reachability analysis is accurate.
An undecontextualizable rule

Initial rule:

\[ \text{Sh}(Y7\sim u,\pi!1),\text{R}(Y48\sim p!1,r!_) \rightarrow \text{Sh}(Y7\sim p,\pi!1),\text{R}(Y48\sim p!1,r!_) \]

Decontextualized rule:

\[ \text{Sh}(Y7\sim u,\pi!1),\text{R}(Y48!1,r!_) \rightarrow \text{Sh}(Y7\sim p,\pi!1),\text{R}(Y48!1,r!_) \]
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)