A Graphical Method For Reducing and Relating Models in Systems Biology [1]

Steven Gay, François Fages, Sylvain Soliman

September 16, 2010

[1] published in Bioinformatics at ECCB'10[Gay et al., 2010]

INRIA

From Models to Metamodels

In Systems Biology, models are built with two contradictory perspectives:

From Models to Metamodels

In Systems Biology, models are built with two contradictory perspectives:

- models for aggregating knowledge: the more concrete the better

From Models to Metamodels

In Systems Biology, models are built with two contradictory perspectives:

- models for aggregating knowledge: the more concrete the better
- models for making predictions: the more abstract the better

From Models to Metamodels

In Systems Biology, models are built with two contradictory perspectives:

- models for aggregating knowledge: the more concrete the better
- models for making predictions: the more abstract the better

These perspectives can be reconciled by organizing models in a **hierarchy of reduction/refinement relations**.

To understand a system is not to know everything about it, but to know abstraction levels that are sufficient for answering questions about the system.

The state of the art on model reduction

Analytical reductions (time/phase decompositions) are too restrictive to be applicable on a large scale.

Published models are not formally related.

Model repositories are still only plain lists of annotated models.

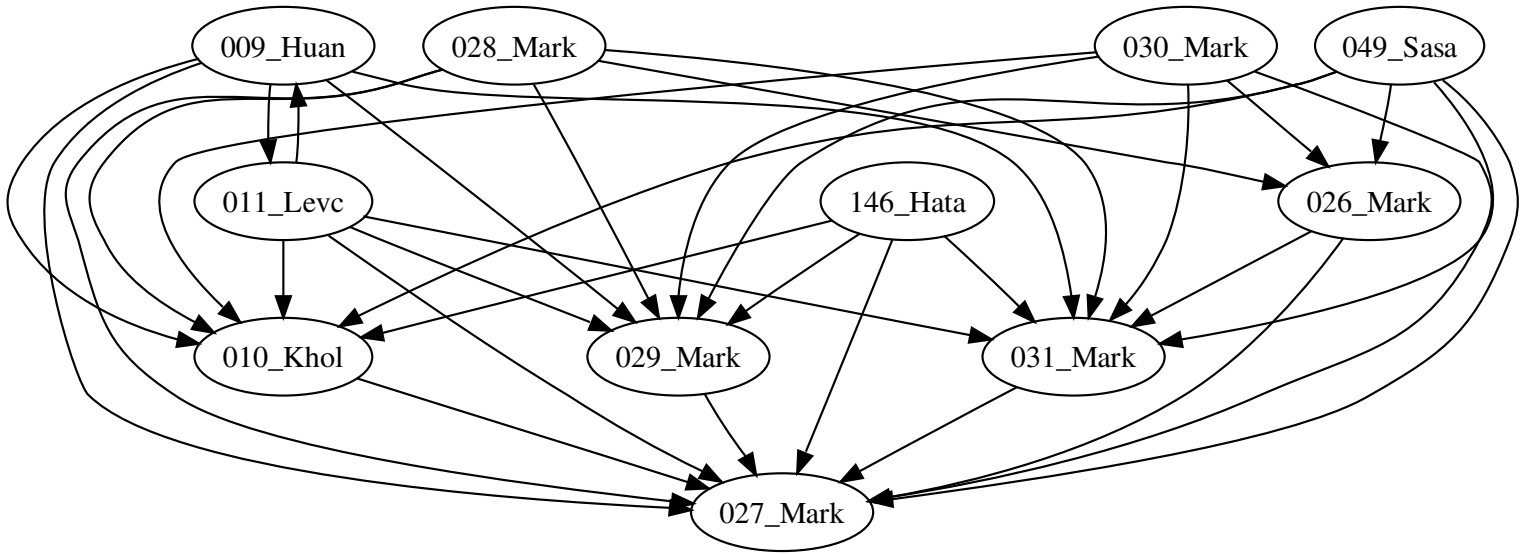
Our contribution

Propose a general, minimalistic formalism for model reduction

Our contribution

Propose a general, minimalistic formalism for model reduction
Compute hierarchies of models

Example (MAPK models in biomodels.net)



INRIA

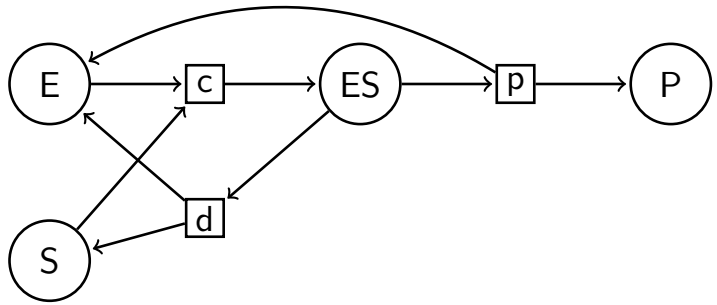
Outline

Reaction Graphs

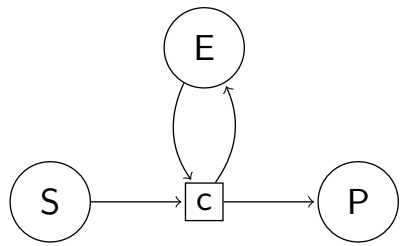
Definition

A *reaction graph* is a triple (S, R, A), with $A \subseteq S \times R \cup R \times S$.
S is the set of *species* of the graph, R is the set of *reactions*.

Example ($E + S \rightleftharpoons ES \rightarrow E + P$)



Example ($E + S \rightarrow E + P$)



INRIA

Model reduction by graph operations

In our setting, we define a model reduction to be a string of the following 4 elementary operations:

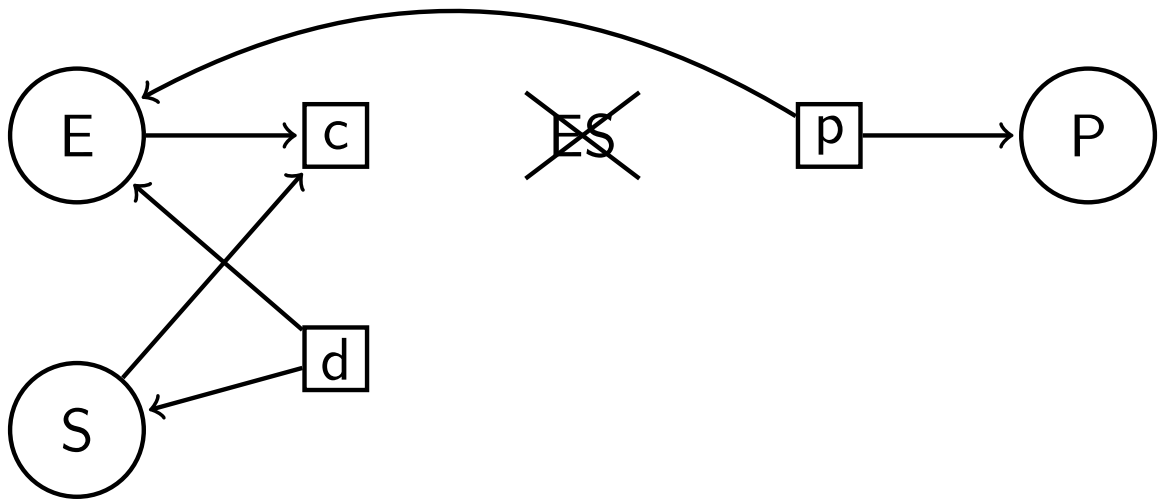
- Species deletion
- Reaction deletion
- Species merging
- Reaction merging

Species Deletion

This removes a species from the model.

- Remove every arc linking the species and any reaction
- Remove the species' node from the graph

Example

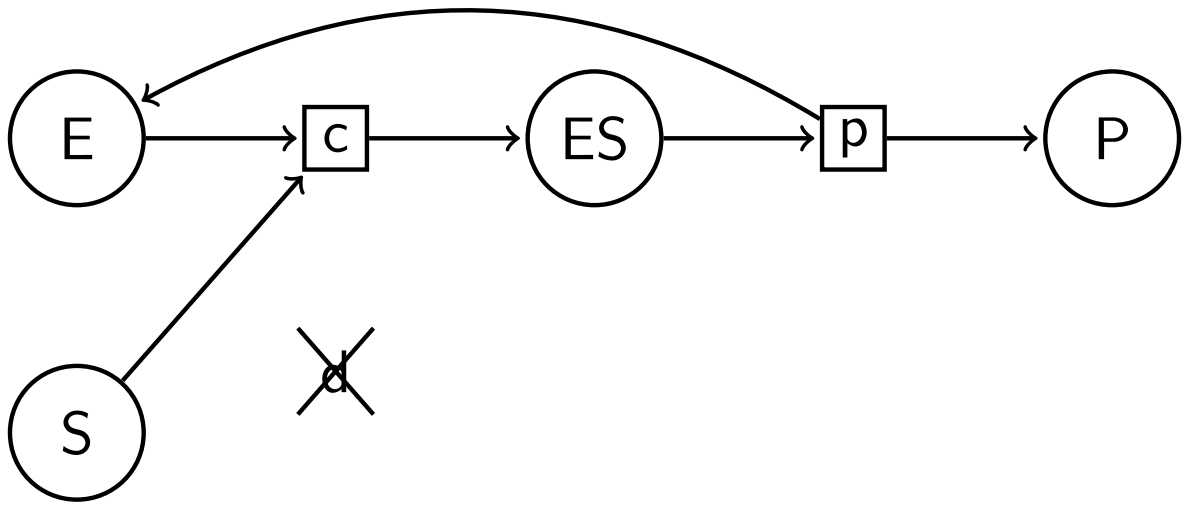


Reaction Deletion

This removes a reaction from the model.

- Remove every arc linking the reaction and a species
- Remove the reaction's node from the graph

Example

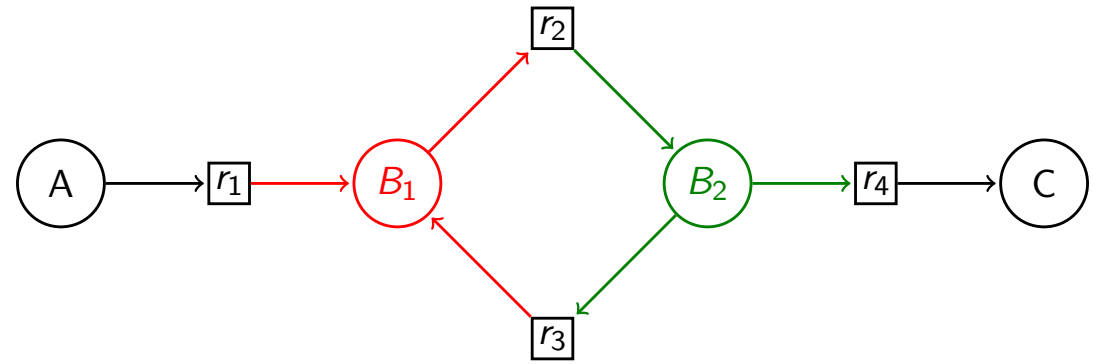


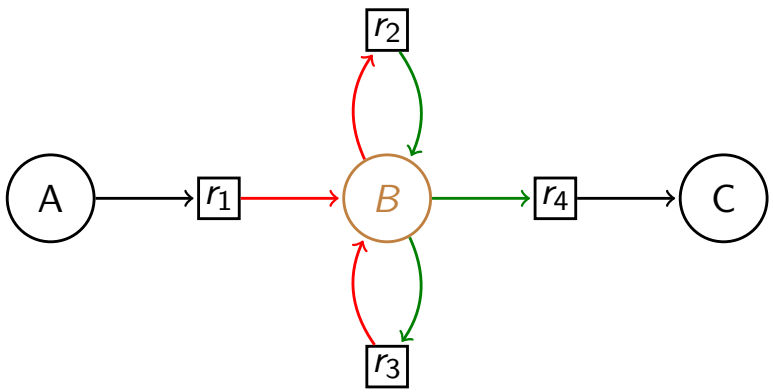
INRIA

Species Merging

This merges several species $S_1 \ldots S_n$ into one:

- Create a new species node S
- For every reaction linked with an S_i, link it with S
- Delete every S_i



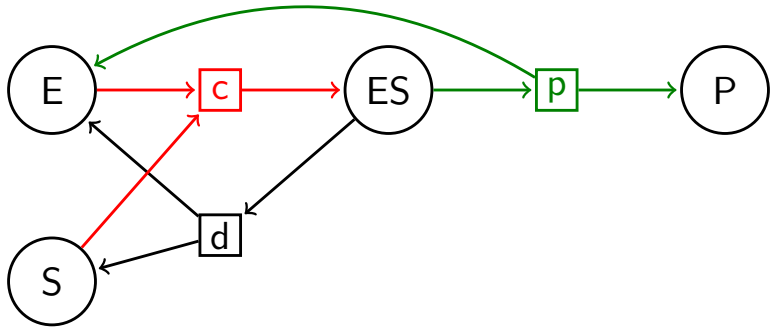


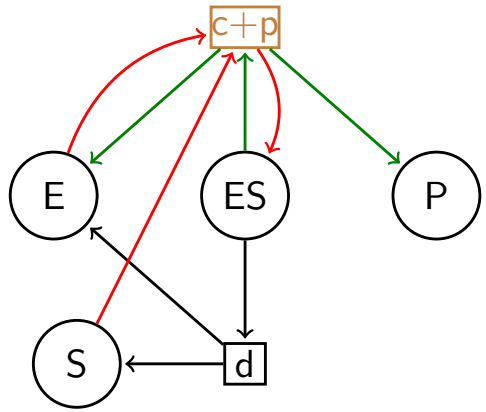
INRIA

Reaction Merging

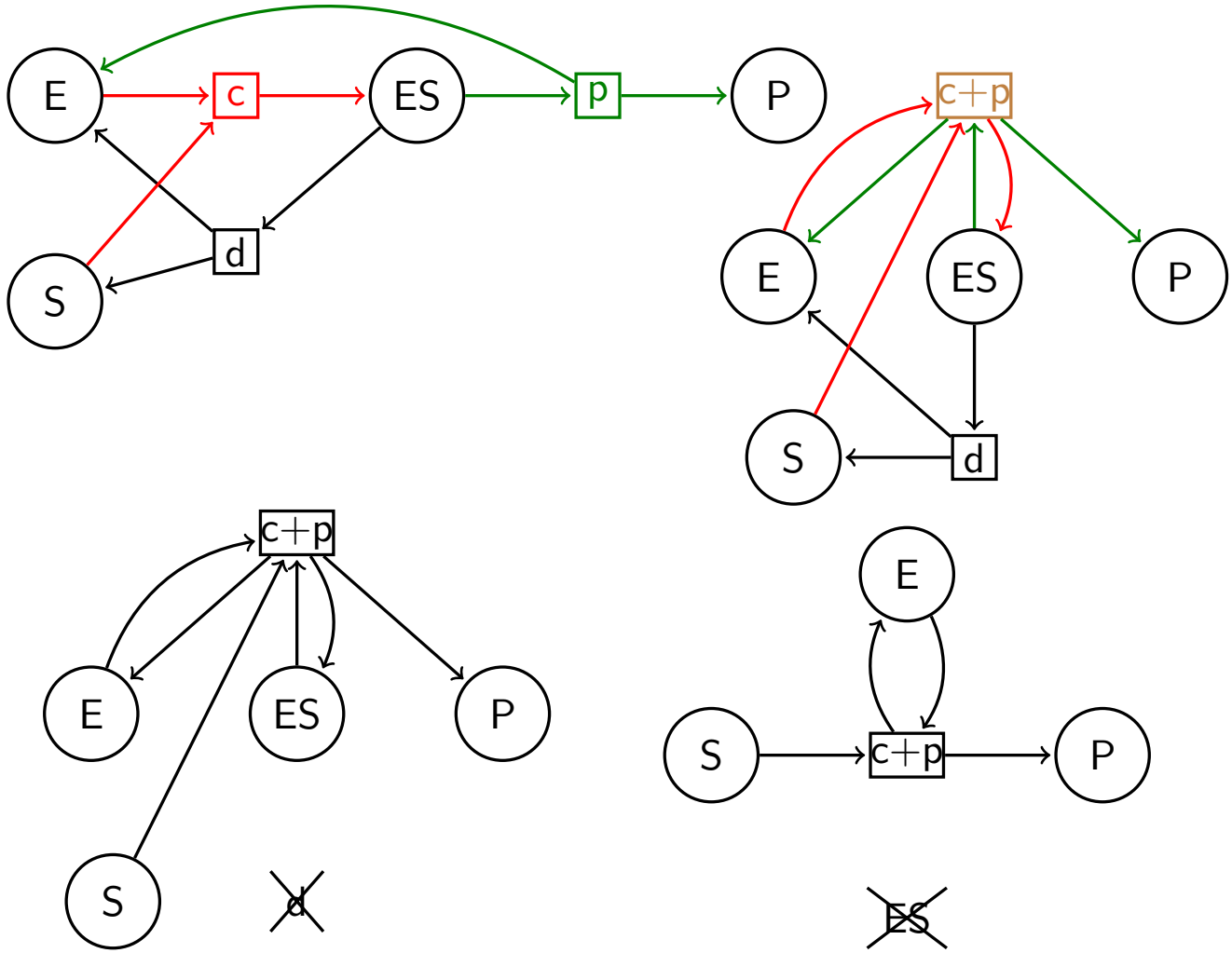
This merges several reactions $R_1 \ldots R_n$ into one:

- Create a new species node R
- For every reaction linked with an R_i, link it with R
- Delete every R_i





The Michaelis-Menten reduction



INRIA

Comparing Graphs with Subgraph Epimorphisms

Definition

A *subgraph morphism* μ from $G = (S, R, A)$ to $G' = (S', R', A')$ is a function $\mu : S_0 \cup R_0 \longrightarrow S' \cup R'$, with $S_0 \subseteq S$ and $R_0 \subseteq R$, such that :

- $\mu(S_0) \subseteq S'$, $\mu(R_0) \subseteq R'$
- $\forall (x, y) \in A \cap (S_0 \times R_0 \cup R_0 \times S_0), (\mu(x), \mu(y)) \in A'$.

Definition

A *subgraph epimorphism* is a subgraph morphism that is surjective (on nodes *and* arcs).

Theorem

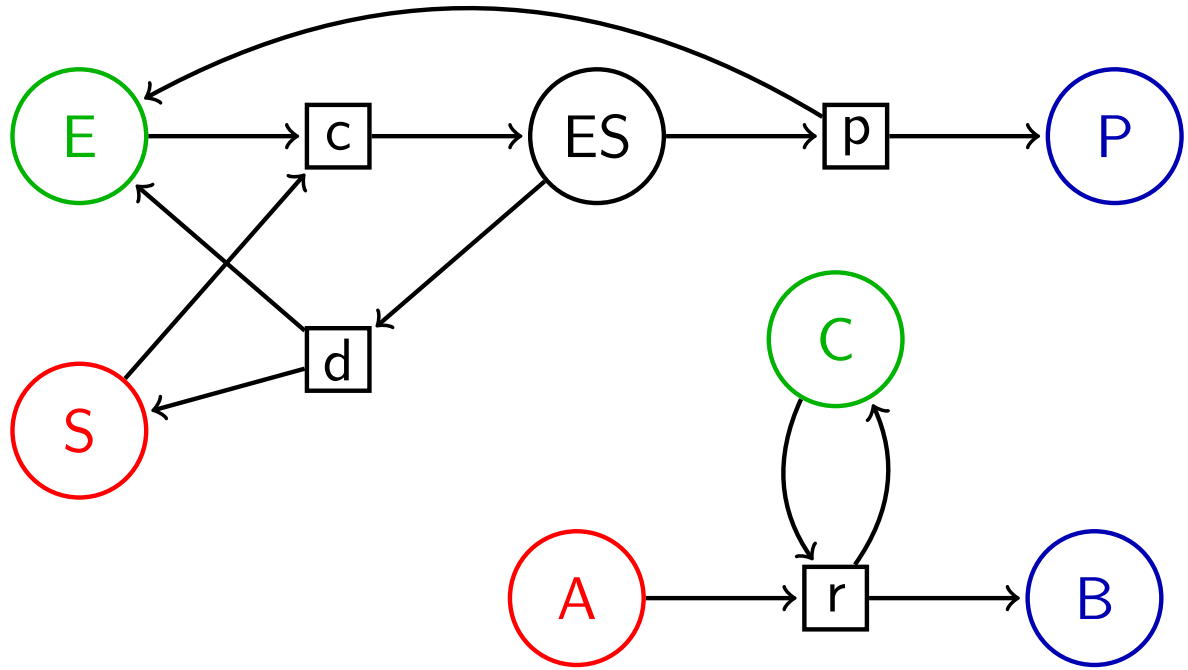
Let G, G' be reaction graphs. There is a reduction from G to G' using species and reaction deletions and mergings iff there is a subgraph epimorphism from G to G'

INRIA

Model Reductions as Subgraph Epimorphisms

Example

$E \to C$
$S \to A$
$P \to B$
$c \to r$
$p \to r$
$d \to \bot$
$\mathrm{ES} \to \bot$



INRIA

The Subgraph Epimorphism Problem

Theorem

The problem of deciding if there is a subgraph epimorphism between two reaction graphs is NP-complete.

Proof (article with Christine Solnon, in preparation):

The problem is trivially in NP.
The completeness is obtained by reduction of the Set Covering Problem. □

Implementation using Constraint Programming

We implemented a program in the GNU-Prolog programming language to solve the Subgraph Epimorphism Problem.

- A finite domain variable is associated to every source node, with domain the nodes of the target graph
- The morphism requirement (arc preservation) is implemented as relational constraints on variables
- The surjectivity constraint is implemented with antecedent variables, one per target arc

Thanks to constraint propagation, the programming language makes an **efficient** enumeration of solutions easy.

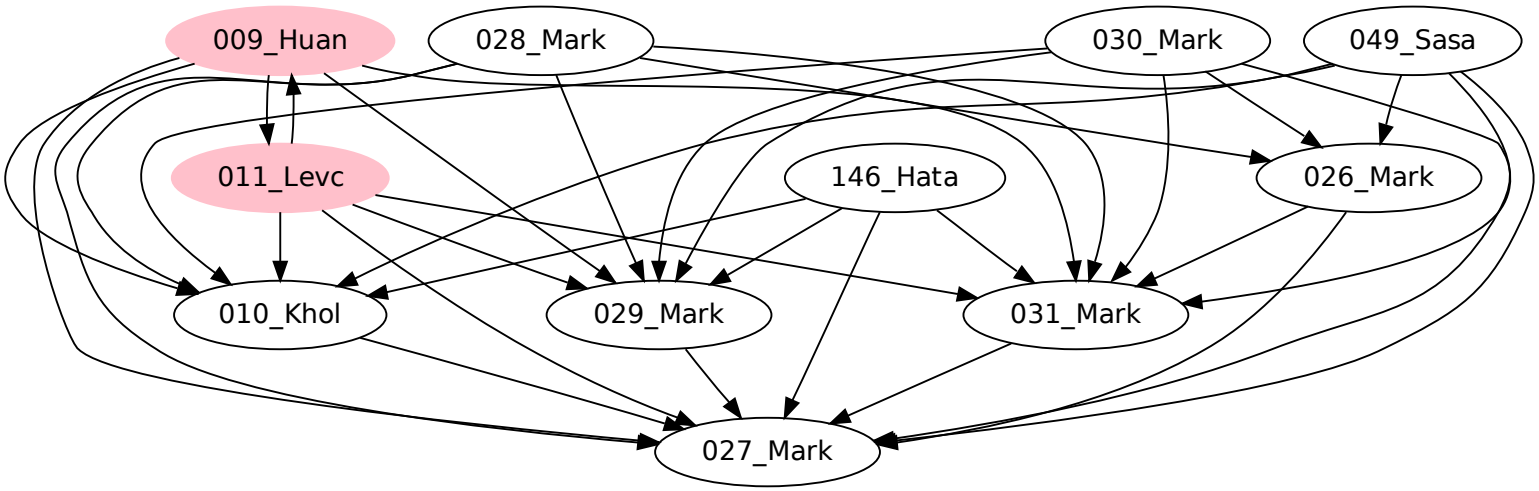
Computing Hierarchies in the Large

To try out the formalism, we extracted four clusters of models from the biomodels.net [le Novère et al., 2006] repository:

- MAPK cascades
- Circadian clock
- Calcium oscillation
- Cell cycle

Every pair of models was compared with a 20 minute timeout ($> 90\%$ of computations took < 5s).

MAPK Hierarchy



This shows the reductions found automatically.

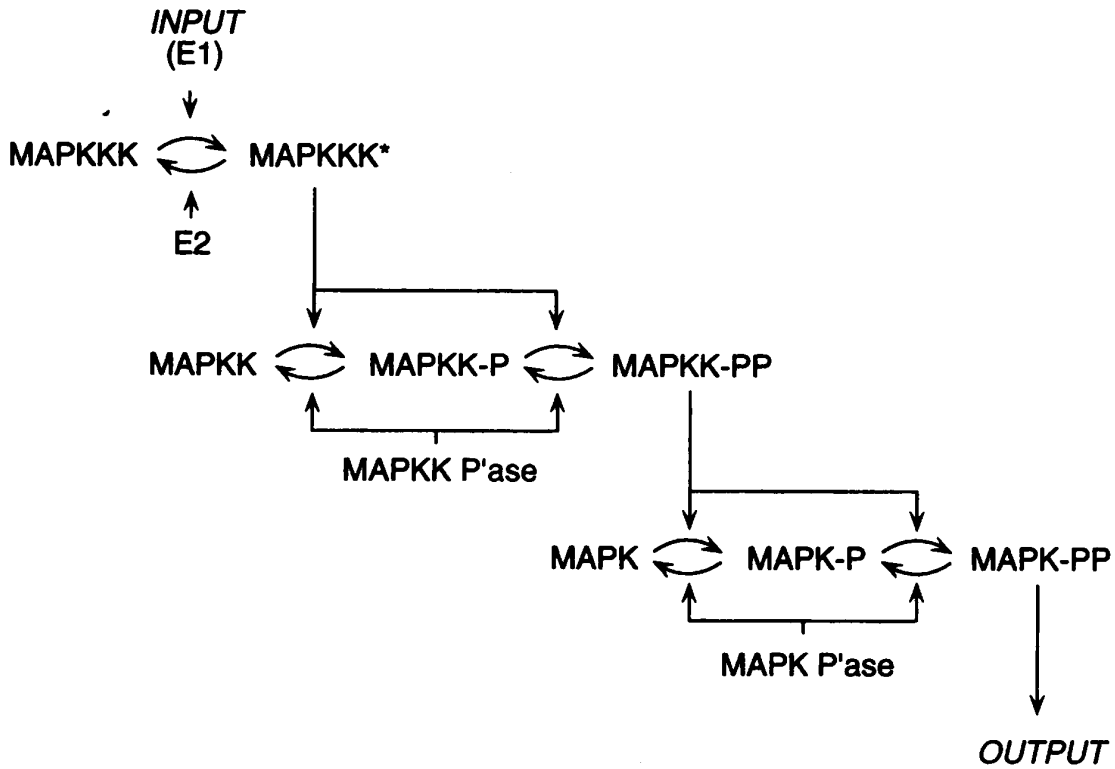
Each node represents a model.

$A \longrightarrow B$ means a reduction from A to B was found.

$A \rightleftharpoons B$ means A and B are isomorphic.

INRIA

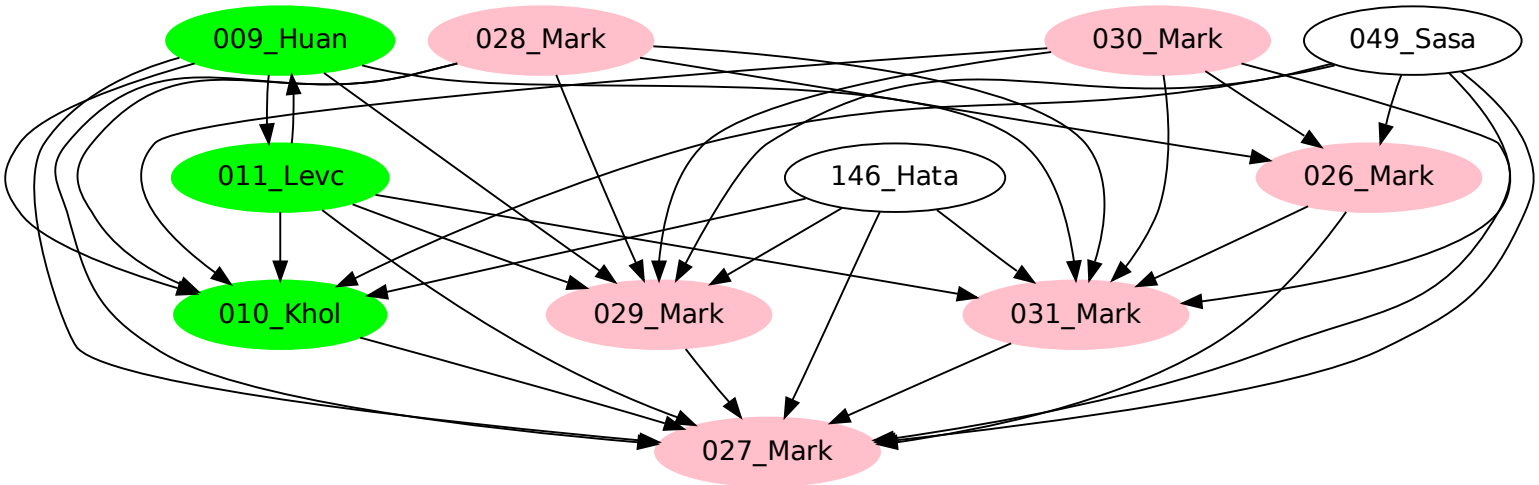
MAPK Models



A MAPK model presented in [Huang and Ferrell, 1996].

INRIA

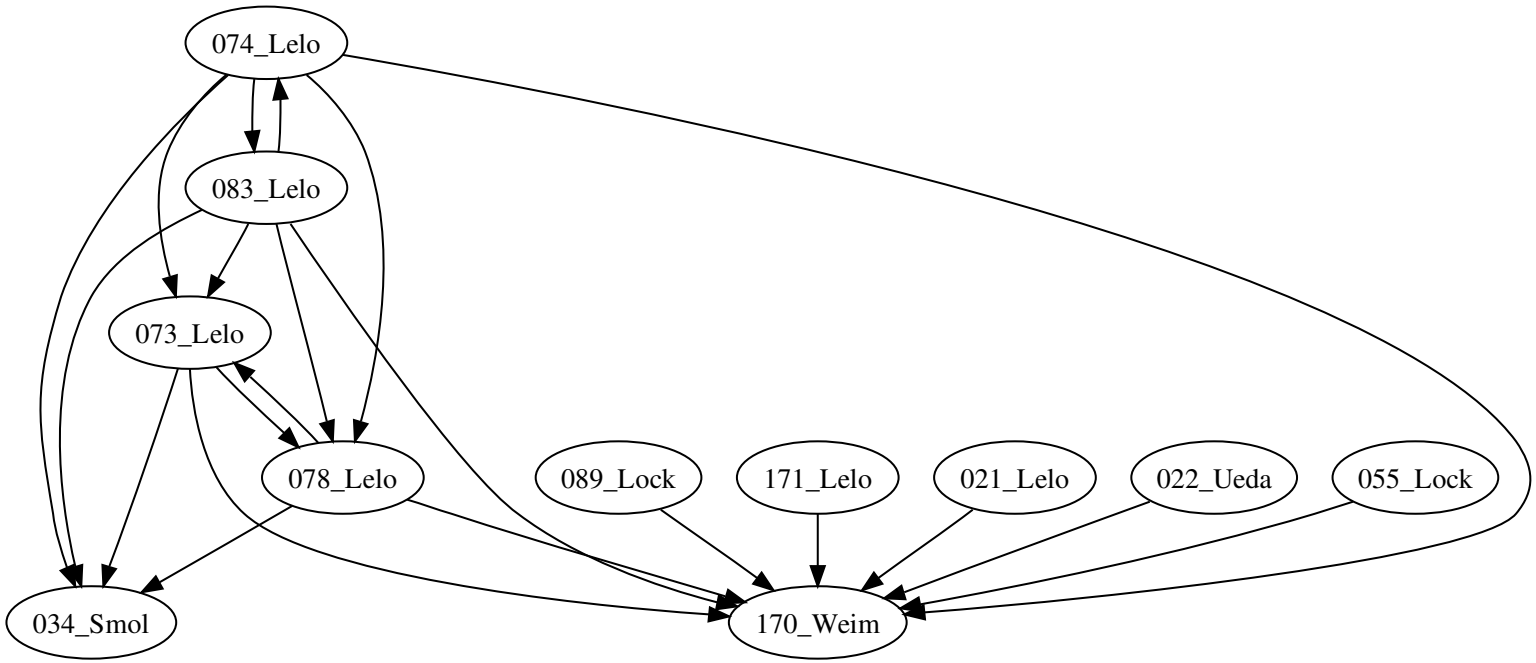
MAPK Hierarchy



Models 009 (Huang 1996), 010 (Kholodenko 2000) and 011 (Levchenko 2000) are 3-level MAPK cascades.
Models 026 to 031 (Markevitch 2004) are 1-level MAPK cascades.
Models 049 (Sasagawa 2005) and 146 (Hatakeyama 2003) are bigger (resp. 216 and 46 reactions), some computations timed out.

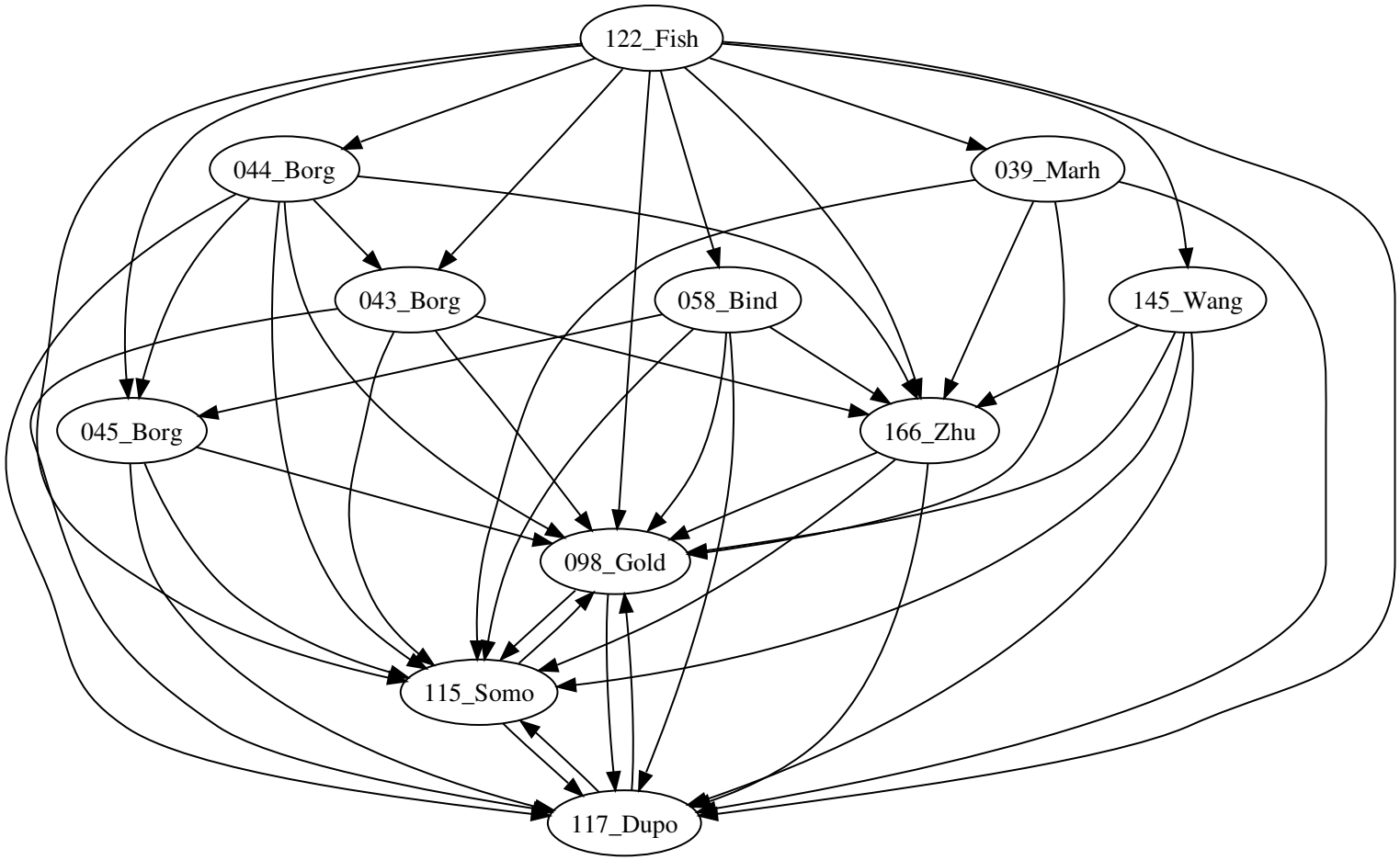
INRIA

Circadian Clock Models Hierarchy



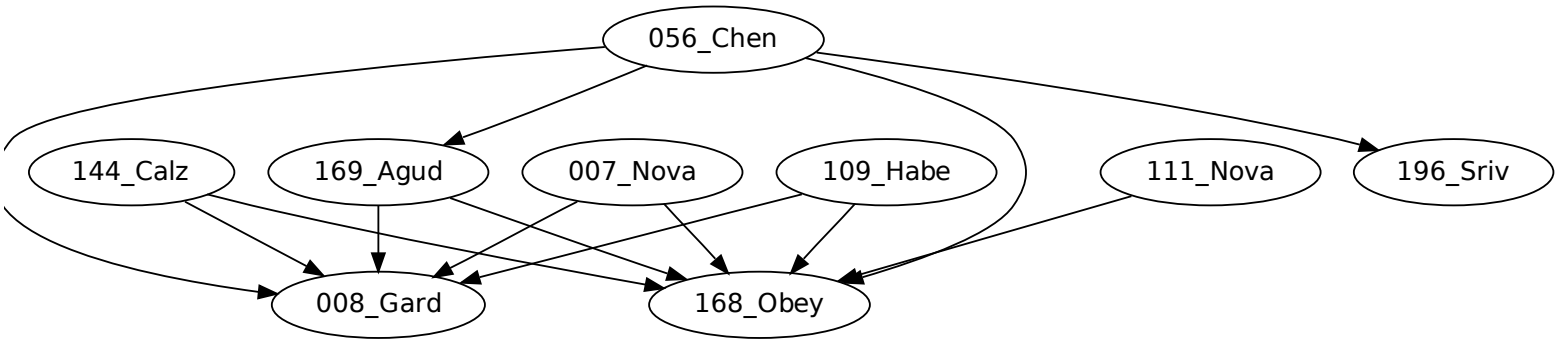
INRIA

Calcium Oscillation Models Hierarchy



INRIA

Cell Cycle Models Hierarchy



Models of the cell cycle in biomodels.net do not represent their structure as reactions.
Typical problems include the use of invariants (matter conservation) and events (cell division), which are not reflected in the reaction graph.

INRIA

Negative Control

We looked for possible reductions between every model of the 4 classes.
Inter-class reductions are false positives, they amount to 9% of the tests.
They are typically big-to-small models reductions.

By removing the smaller models from the pool, only 1.2% of the tests are positives remain.

Conclusion

The model reduction method we presented is stricly structural. The four graph reduction operations are simple, but precise enough to capture model reductions in biomodels.net.

As a future work, linking the operations with mathematical reductions is the obvious path:

- species deletions with species in excess
- reaction deletions with slow reverse reactions
- species mergings with fast equilibria (QSSA)
- reaction mergings with limiting reactions

Thank you for your attention!

Questions?

INRIA


Gay, S., Soliman, S., and Fages, F. (2010).
A graphical method for reducing and relating models in systems biology.
Bioinformatics, special issue ECCB'10.


Huang, C.-Y. and Ferrell, Jr., J. E. (1996).
Ultrasensitivity in the mitogen-activated protein kinase cascade.
PNAS, 93(19):10078–10083.


le Novère, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li, L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J. L., and Hucka, M. (2006).
BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems.
Nucleic Acid Research, 1(34):D689–D691.

INRIA