

Analysis and repair of whole genome bacterial metabolic models for Synthetic Biology

Monika Heiner^{1,2} & David Gilbert²

¹ Brandenburg Technical University (BTU), Cottbus, Germany

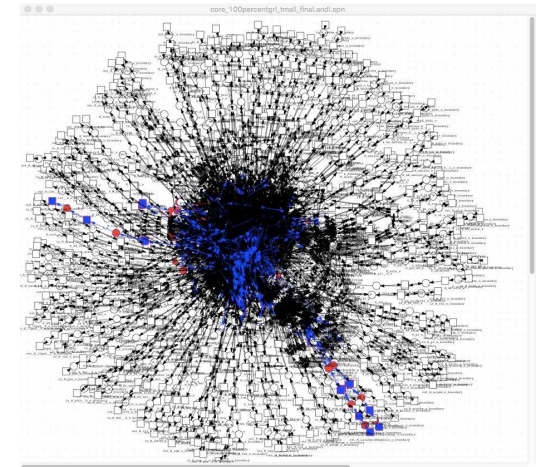
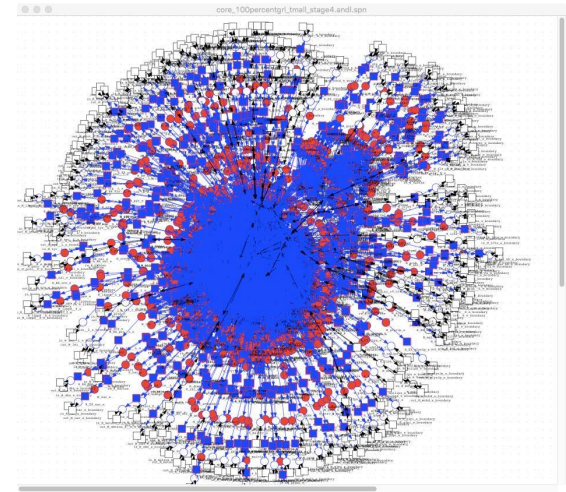
² Brunel University London, UK,
Synthetic Biology Theme & Department of Computer Science

danke - 感謝 - thanks



Outline

- Brunel University London:
bacterial engineering / Synthetic Biology
- Whole genome metabolic models
 - engineering design templates
- Need for 'correct' initial template description
 - well behaved (dynamic behaviour)
- Based on (badly behaved) public domain models
- Structure based correction of initial models
 - graph analysis, graph editing,
dynamic simulation, model checking
- Initial Brunel core model



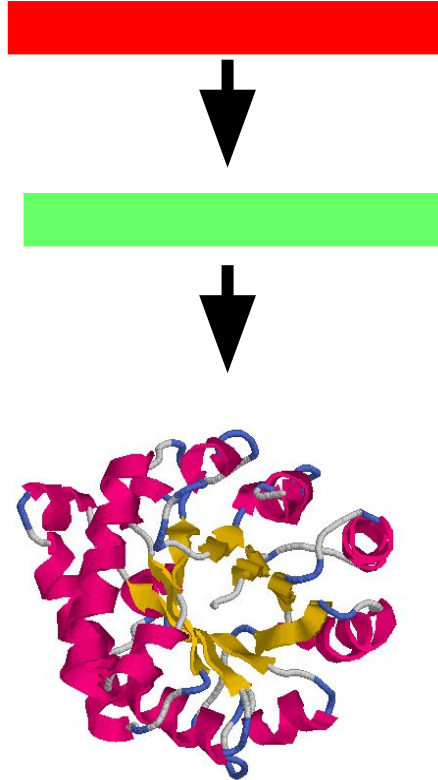
DNA

"gene"

mRNA

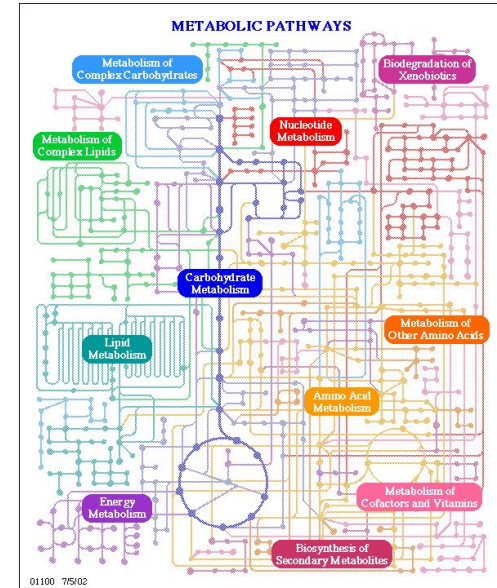
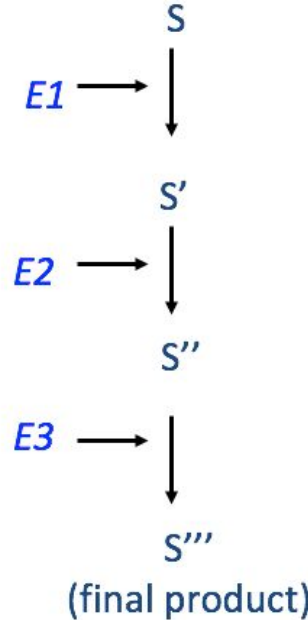
Protein
sequence

Folded
Protein

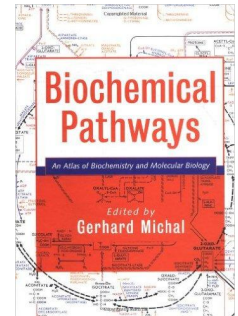
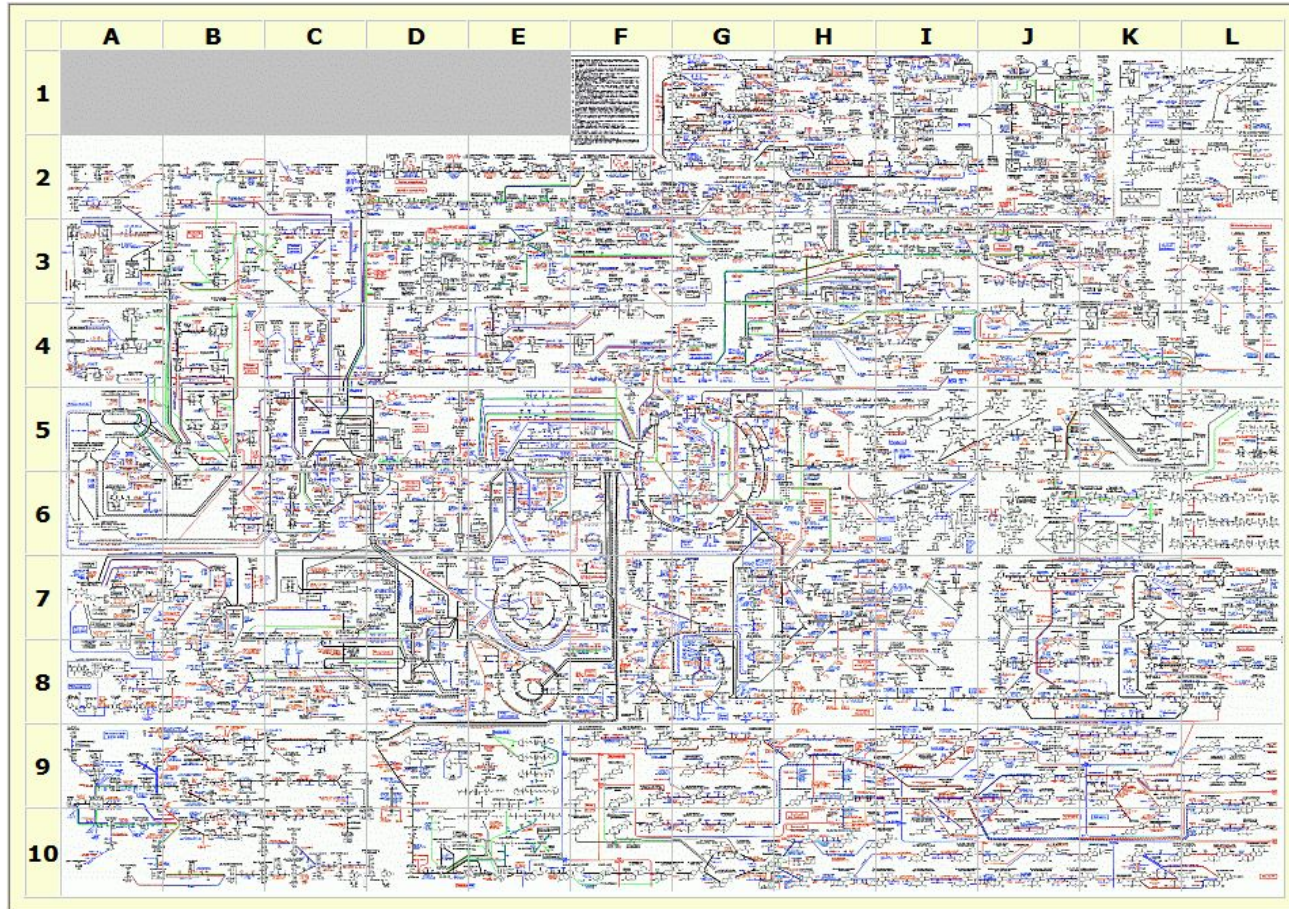


From Genes to Systems

(initial substrate)



Metabolic Pathways



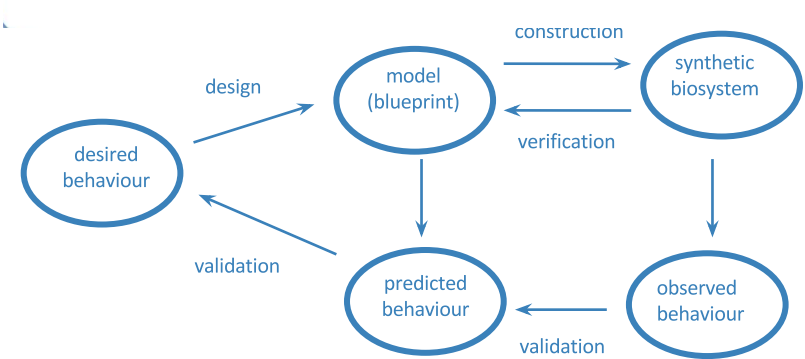
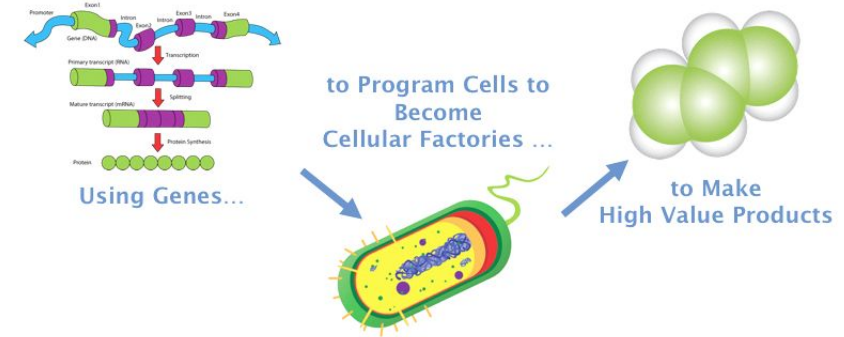
Synthetic Biology / Bacterial Engineering

- **Make a new one**

- system, or
- product

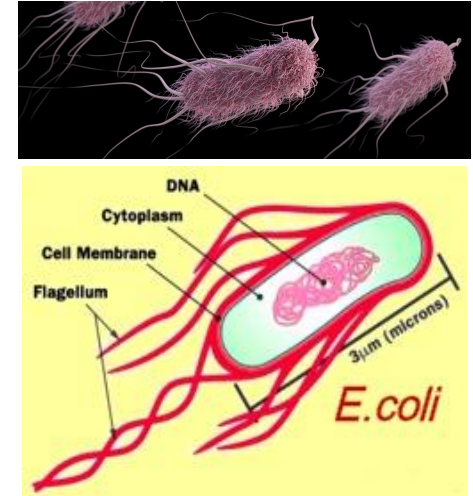
- **Synthetic Biology**

- the structured engineering of biological systems for useful purposes

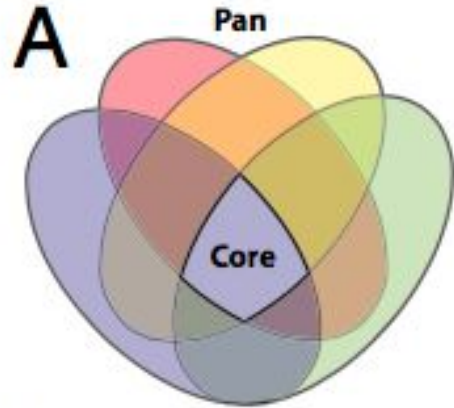


The PROJECT: Design Methods for Bacterial Engineering

- **focus on *E.coli*** [so far]
 - gram-negative, anaerobic, rod-shaped bacterium commonly found in lower intestine of warm-blooded organisms
 - > 4k protein coding genes
 - public domain model collection [Monk et al. 2013]
 - **designed for FBA -> steady state analysis**
- **... to develop computational techniques**
 - **dynamic simulation -> transient behaviour analysis**
- **... to build the Brunel Core Model**
 - based on gene set from Nigel Saunder's group



Monk Metabolic Models

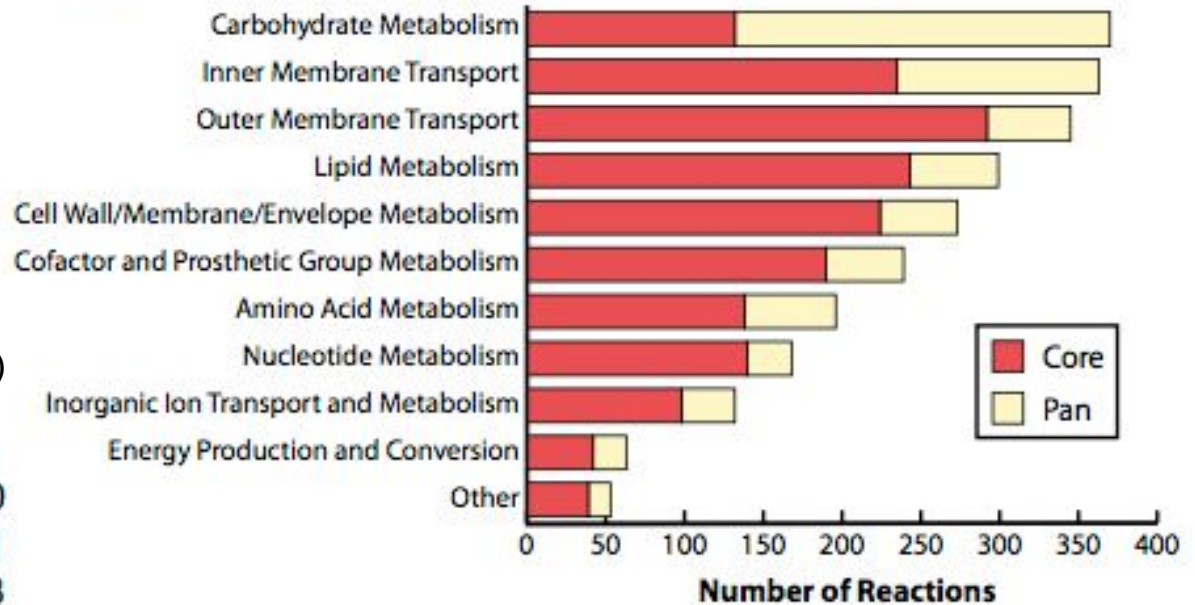


For 55 unique *E. coli* strains *)

	Core:	Pan:
Genes:	965	1,460
Reactions:	1,773	2,501
Metabolites:	1,665	2,043

B

Reaction Distribution by Subsystem

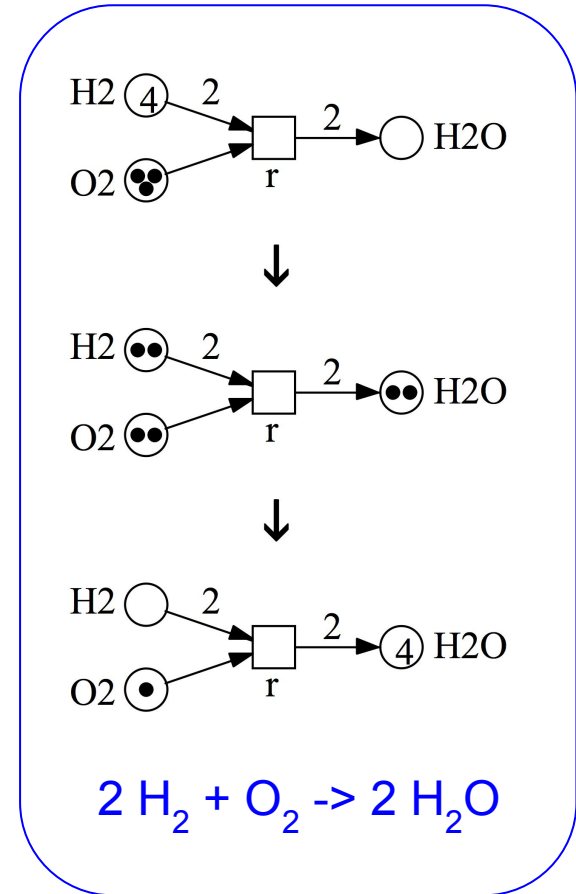


*) 47 *E. coli*, 8 *Shigella*

[Monk 2013]

Biological Models

- reaction/metabolite graphs
 - bipartite graphs → Petri nets
- stoichiometry / arc weights
- no kinetic rates
 - assume mass action, rate=1, if any
- boundary conditions
- model structure
 - cytoplasm, periplasm, external, boundary
- SBML (Systems Biology Markup Language)
 - → Petri nets



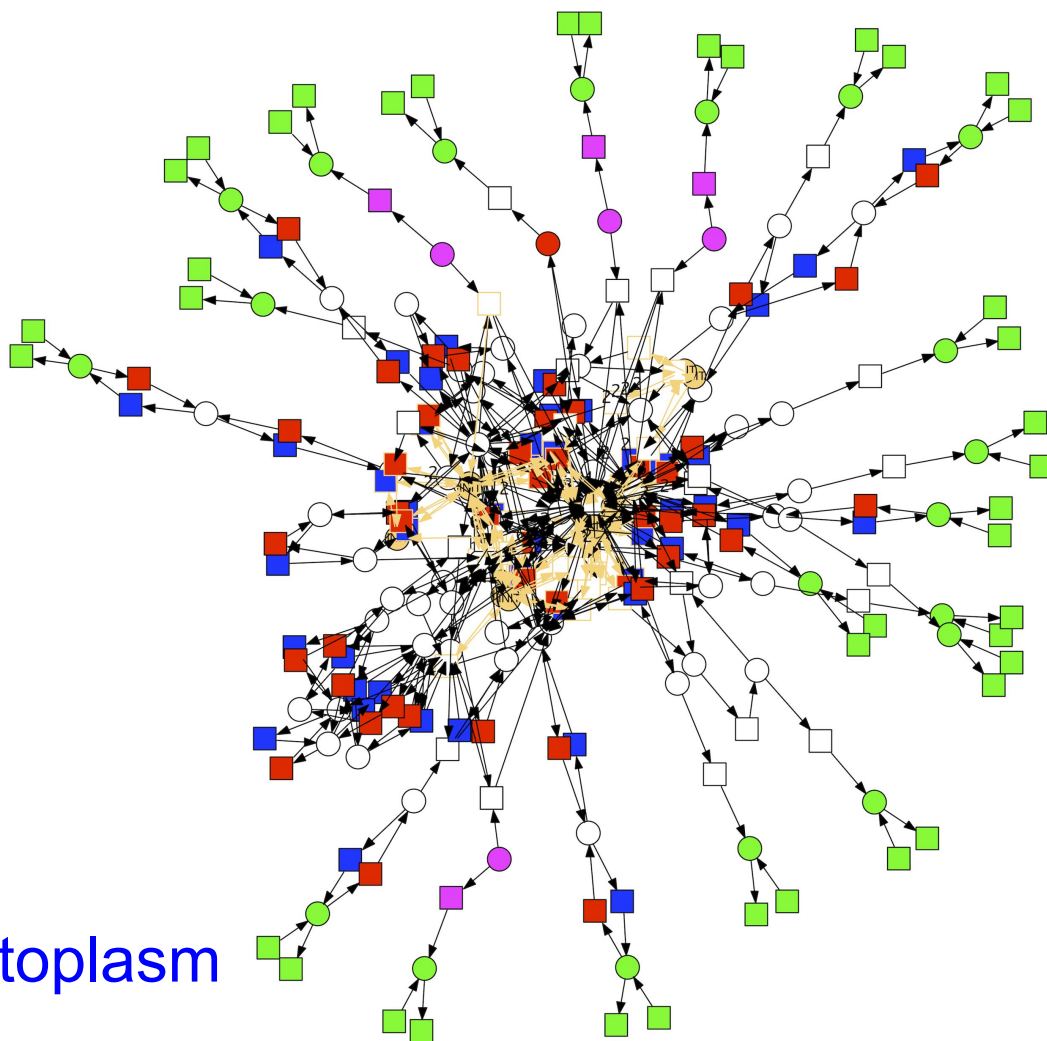
Example E. coli core

[Orth 2010]

model structure:

- cytoplasm,
- periplasm,
- external,
- boundary

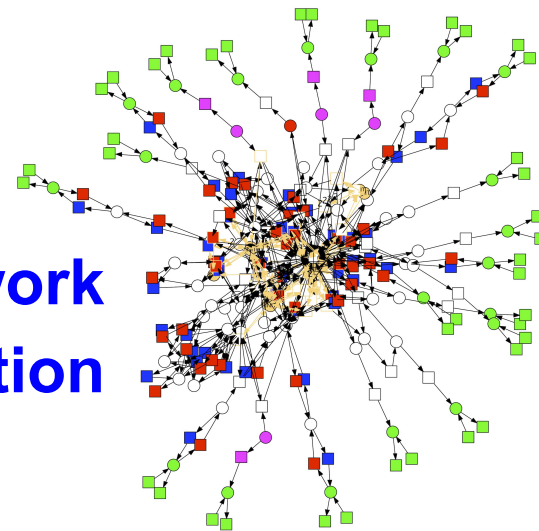
in/out flow through cytoplasm



Assumption

We postulate that a **'good' metabolic network** is one in which **every metabolite and reaction** is (at least)

- weakly live (i.e. exhibits dynamic behaviour) at some point, and
- has a non-zero steady state.

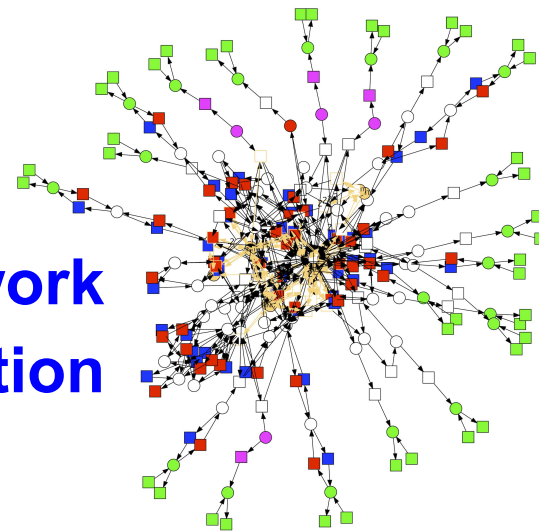


Assumption

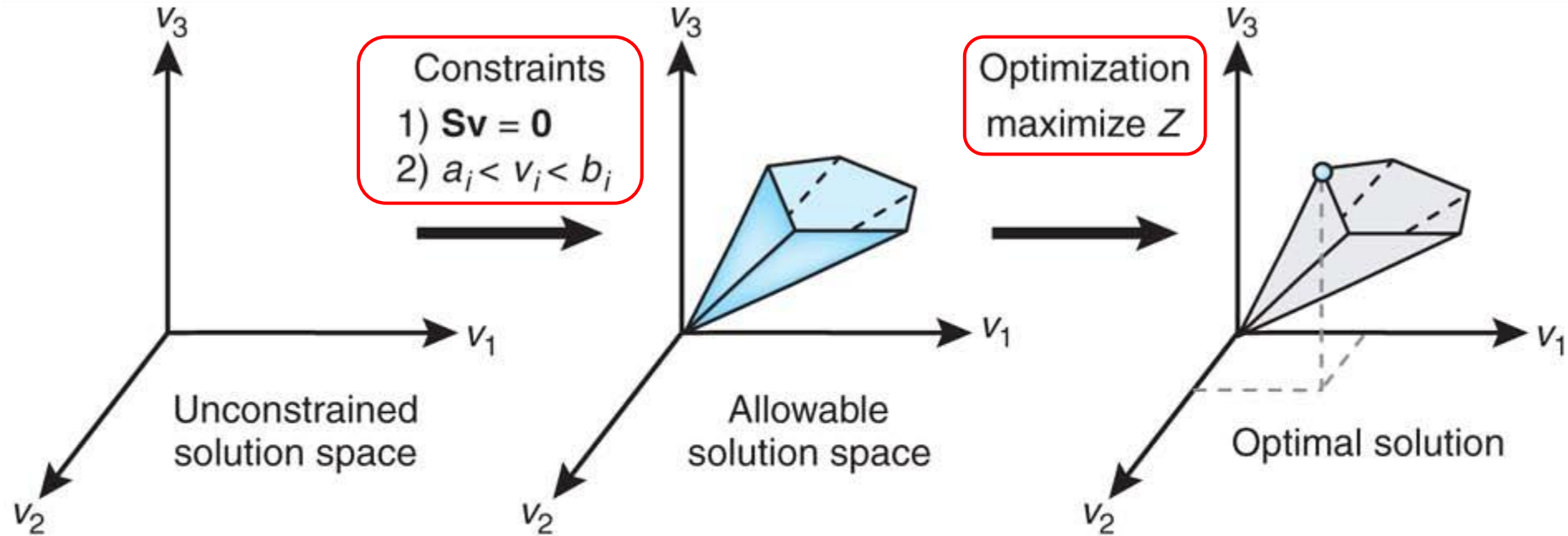
We postulate that a **'good' metabolic network** is one in which **every metabolite and reaction** is (at least)

- weakly live (i.e. exhibits dynamic behaviour) at some point, and
- has a non-zero steady state.

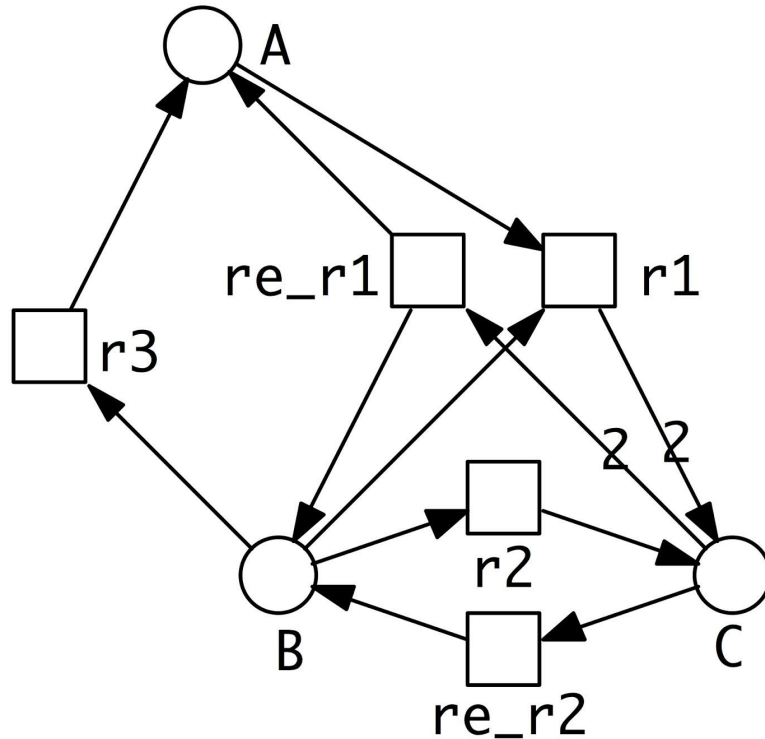
SOUNDS EASY, BUT ISN'T, BECAUSE . . .



Flux Balance Analysis (FBA)

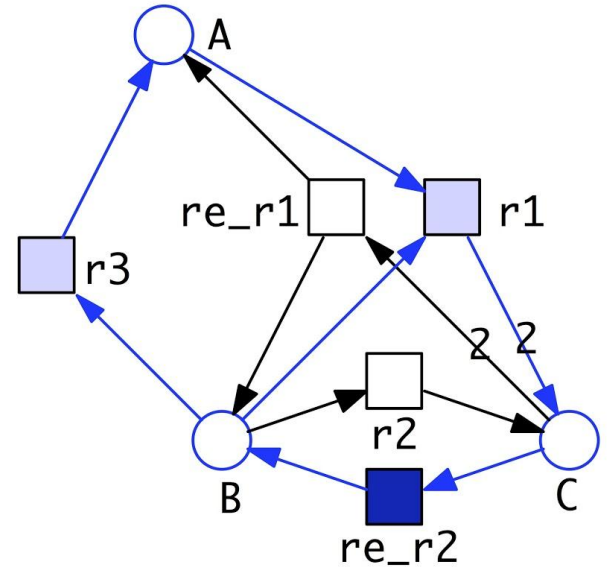
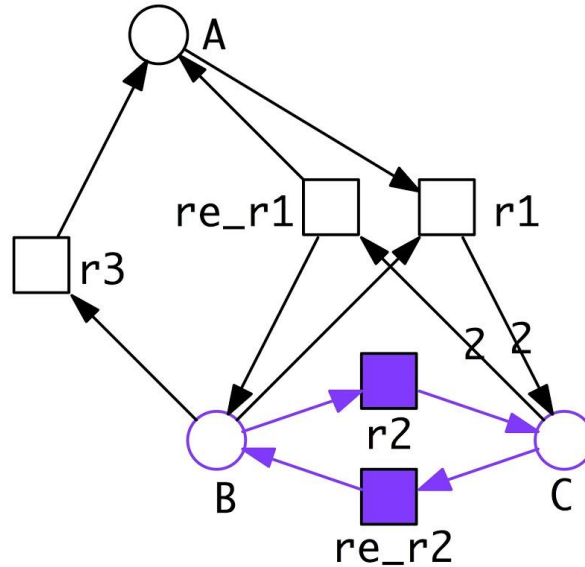
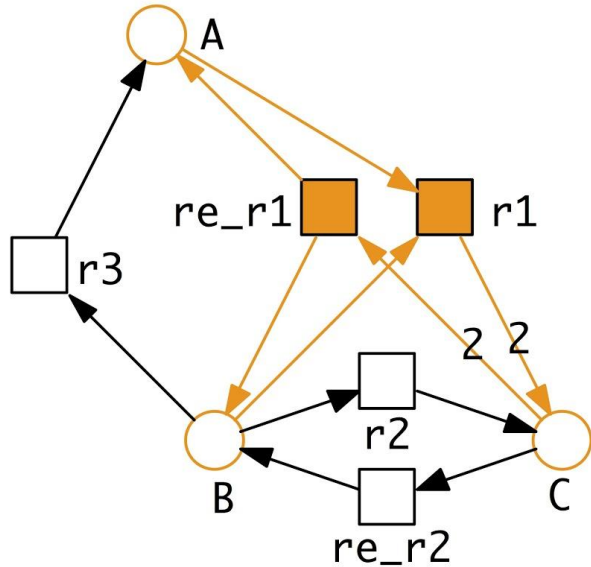


Example



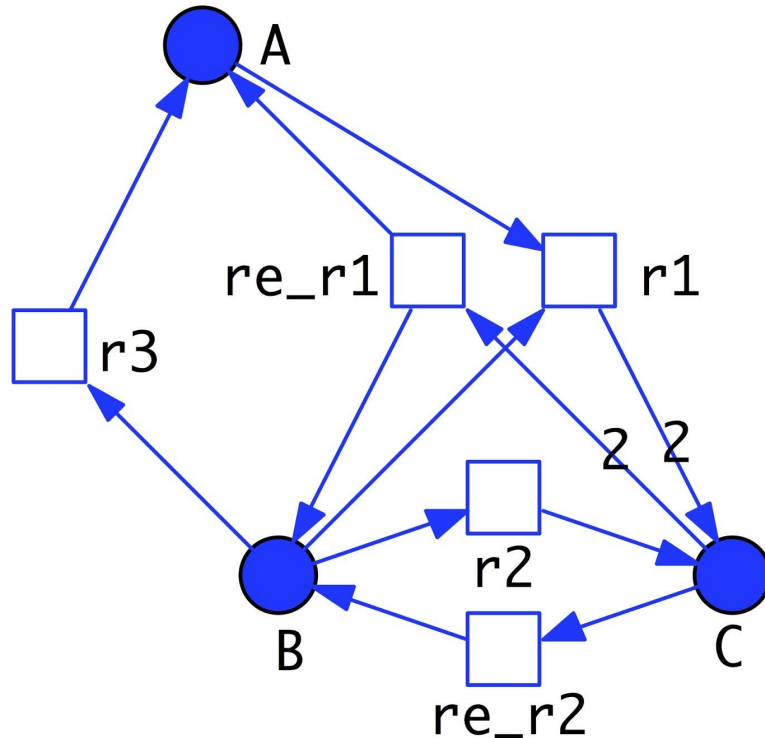
$A+B$	\leftrightarrow	$2C$
B	\leftrightarrow	C
B	\rightarrow	A

Example - T-invariants



-> covered with T-invariants (CTI)

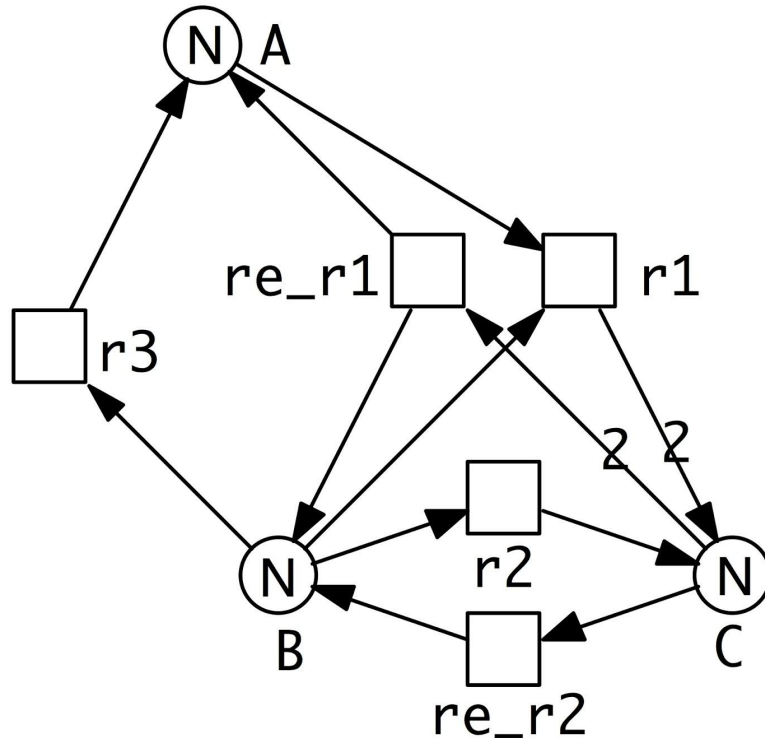
Example - P-invariants



$A+B$	\leftrightarrow	$2C$
B	\leftrightarrow	C
B	\rightarrow	A

\rightarrow covered with
P-invariants
(CPI)

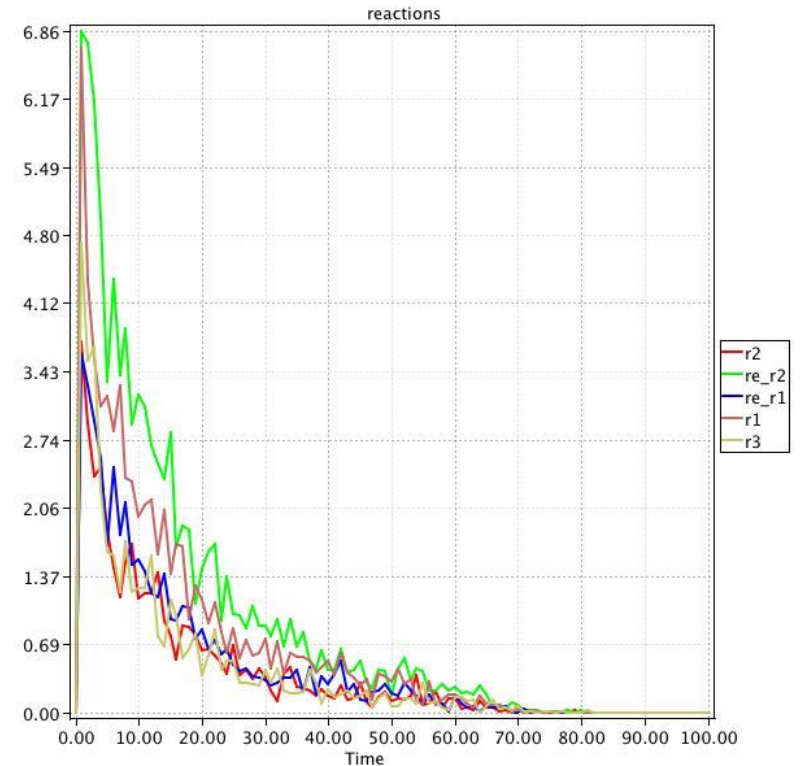
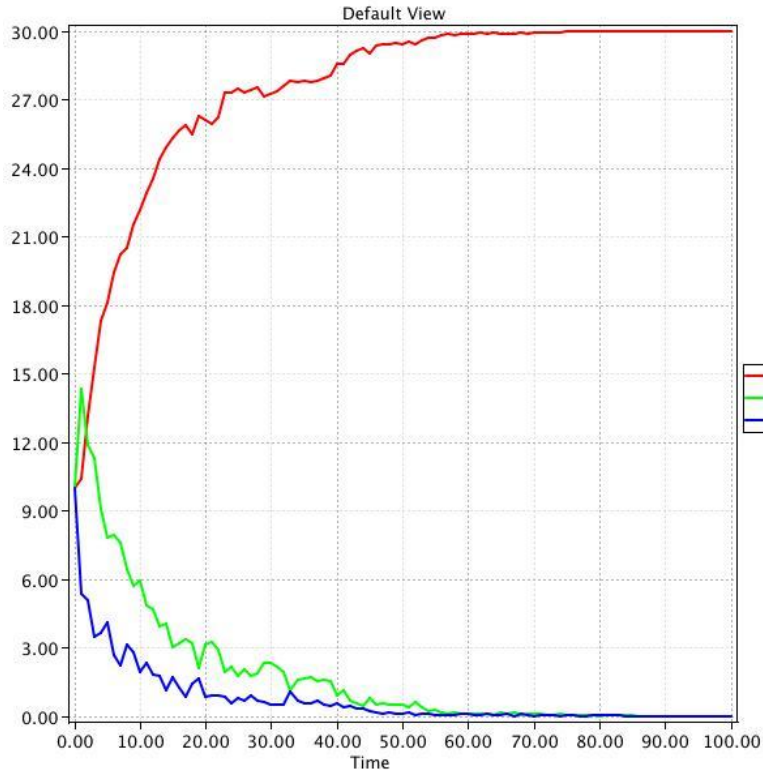
Example - Initialisation



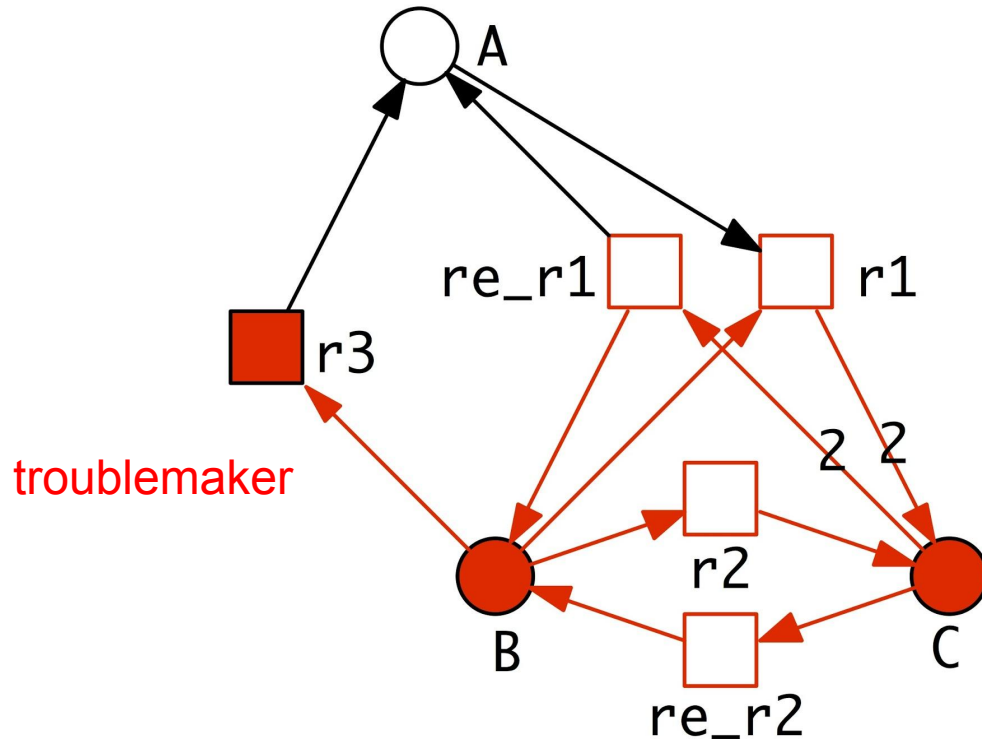
$A+B$	\leftrightarrow	$2C$
B	\leftrightarrow	C
B	\rightarrow	A

Const $N = 1, 5, 10, 50, 100, \dots$

Example - Simulation Results (N=10)

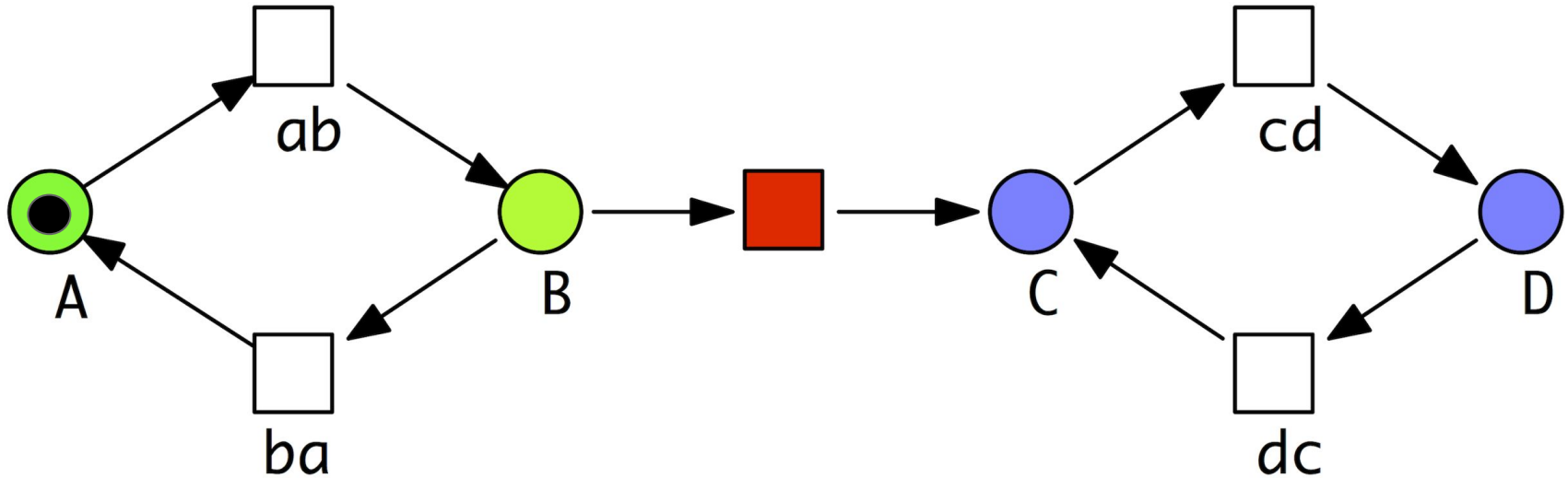


Example - Bad Siphon



$A+B$	\leftrightarrow	$2C$
B	\leftrightarrow	C
B	\rightarrow	A

Simple Siphon / Trap

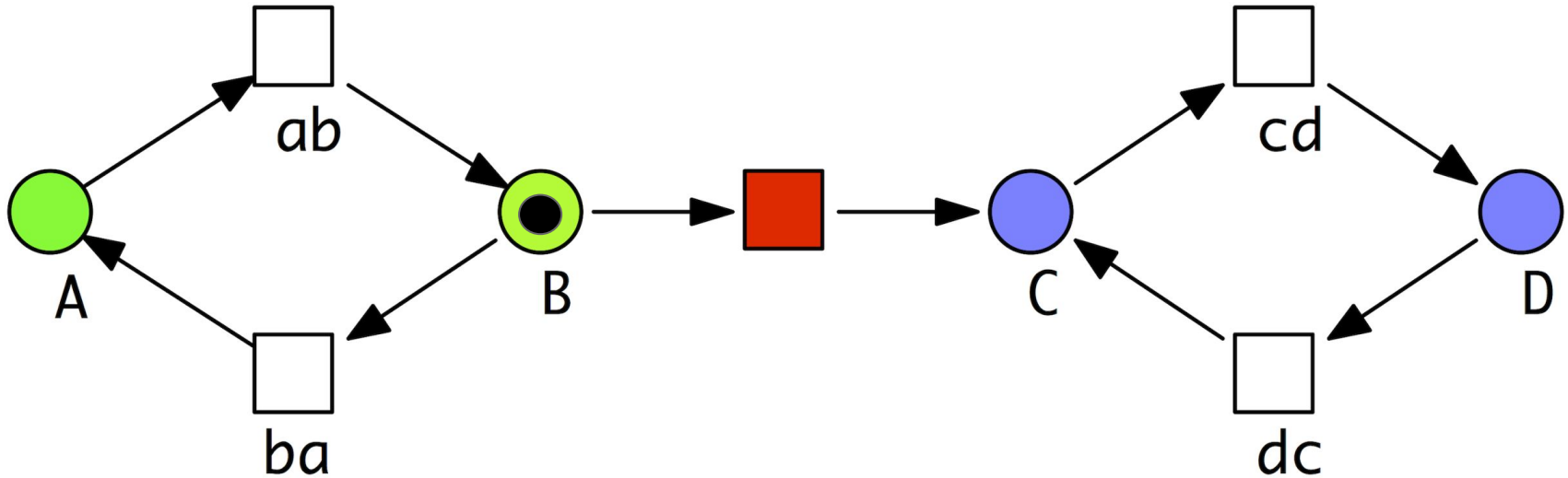


siphon places

troublemaker transition

trap places

Simple Siphon / Trap

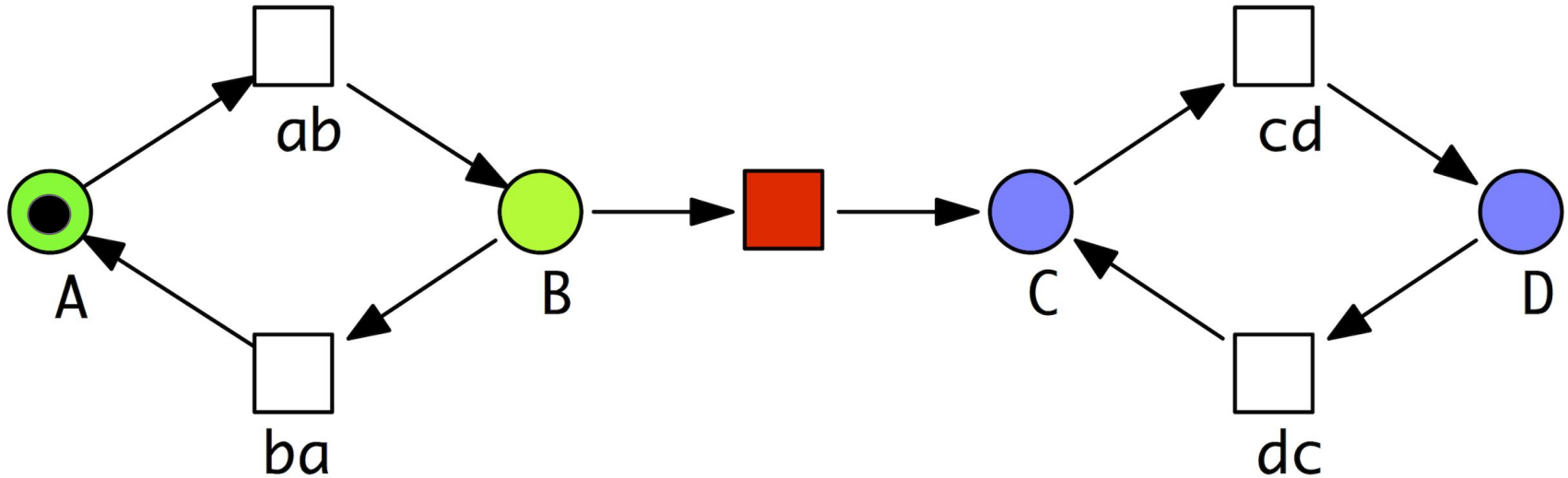


siphon places

troublemaker transition

trap places

Simple Siphon / Trap

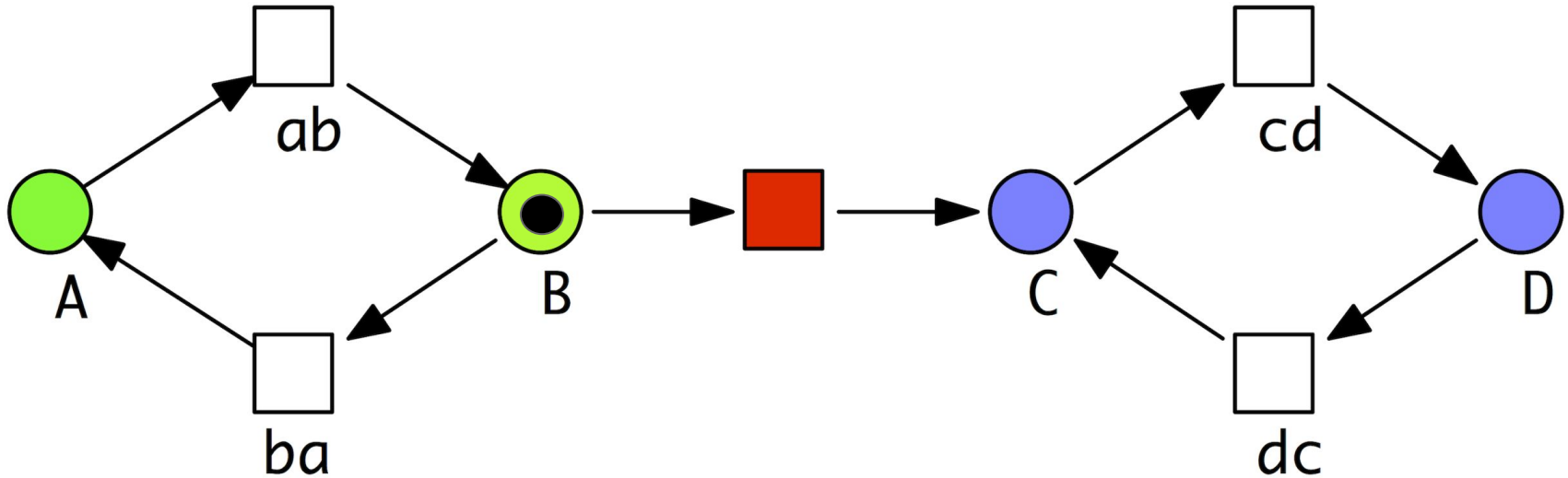


siphon places

troublemaker transition

trap places

Simple Siphon / Trap

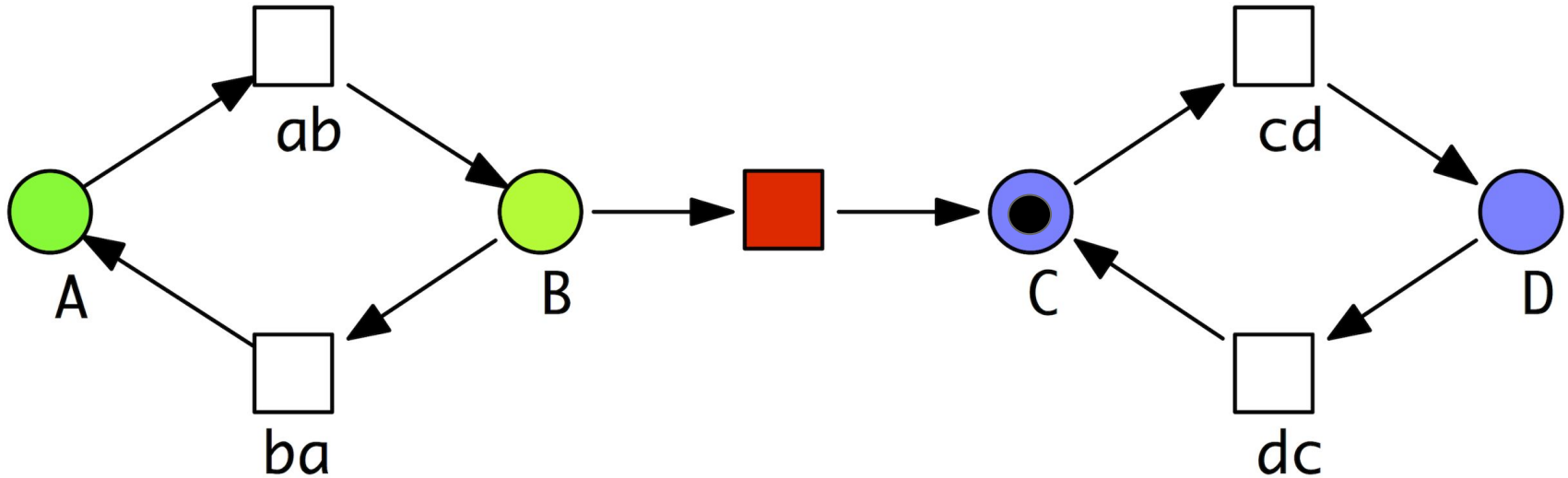


siphon places

troublemaker transition

trap places

Simple Siphon / Trap

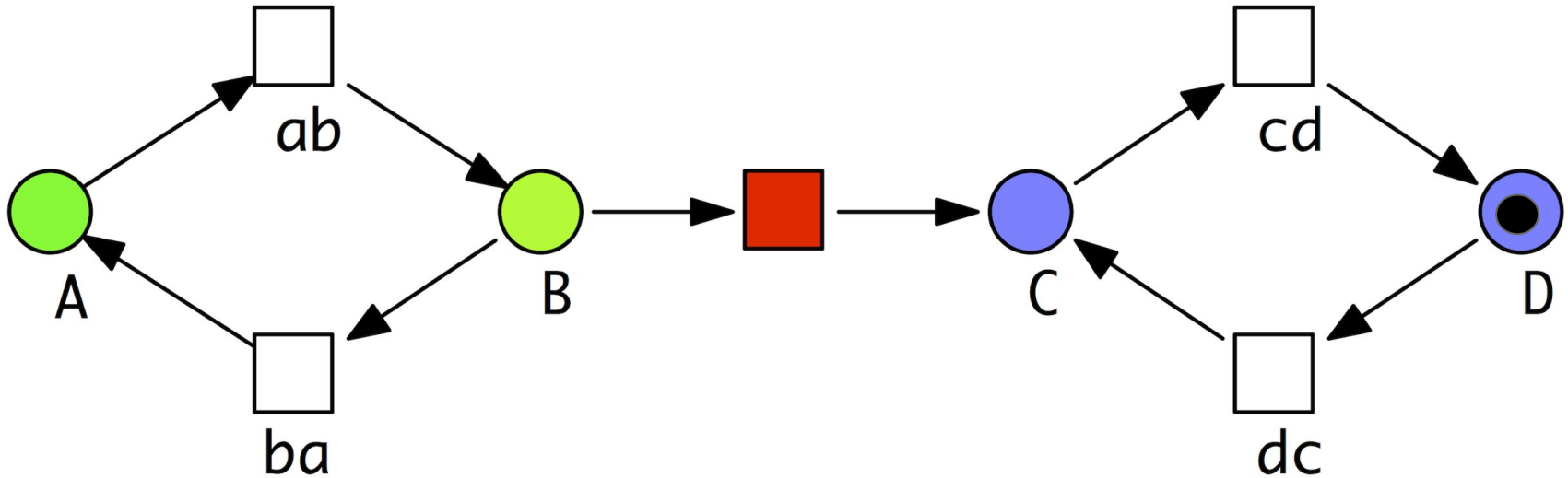


siphon places

troublemaker transition

trap places

Simple Siphon / Trap

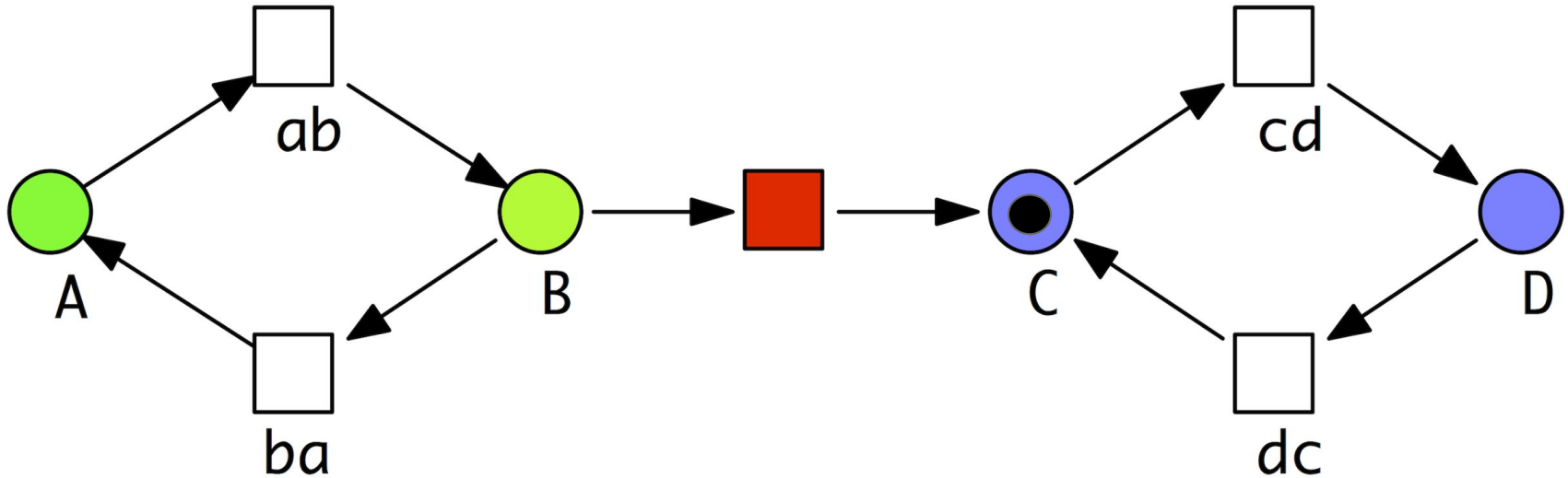


siphon places

troublemaker transition

trap places

Simple Siphon / Trap

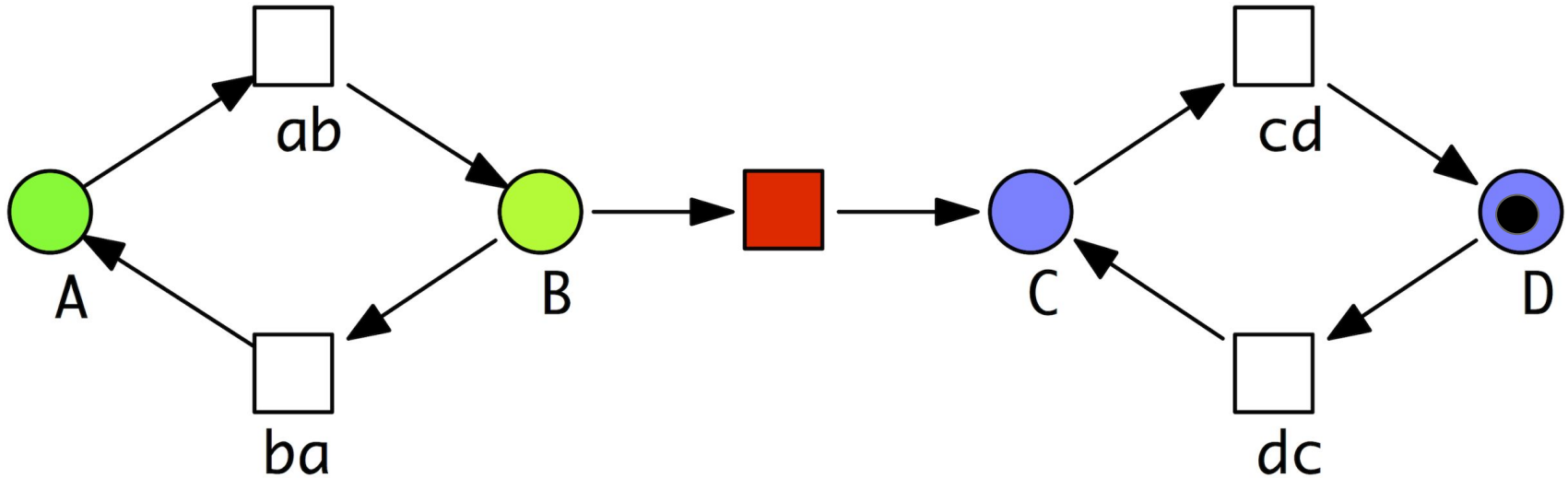


siphon places

troublemaker transition

trap places

Simple Siphon / Trap

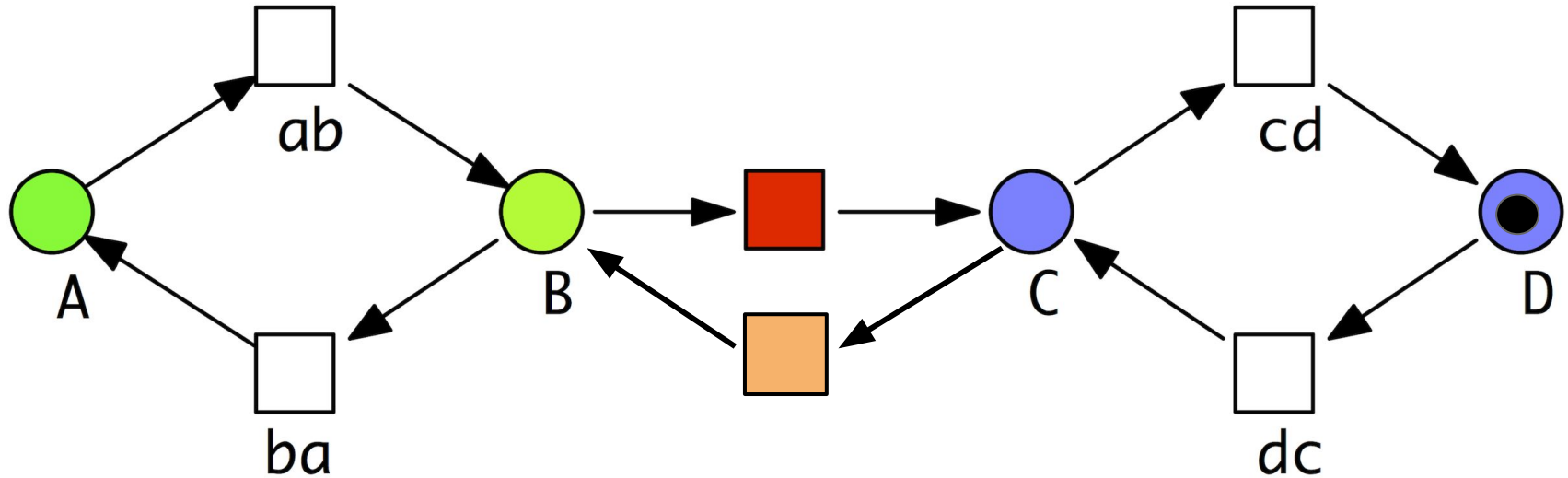


siphon places

troublemaker transition

trap places

Simple Siphon / Trap



siphon places

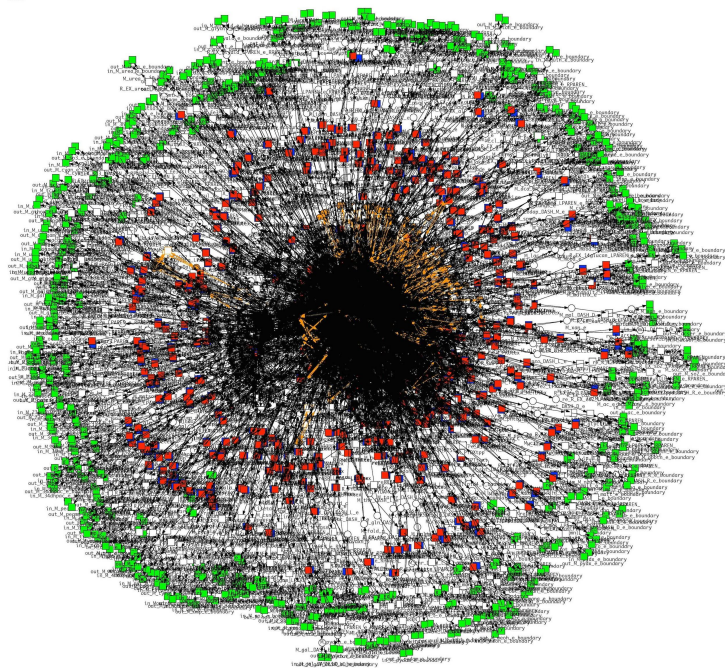
troublemaker transition

trap places

repair transition

Computational Challenges (1)

- large size models
- example sizes
 - reactions > 4k
 - metabolites > 2k
 - connected by > 13k arcs
 - metabolite connectivity: 2-1200

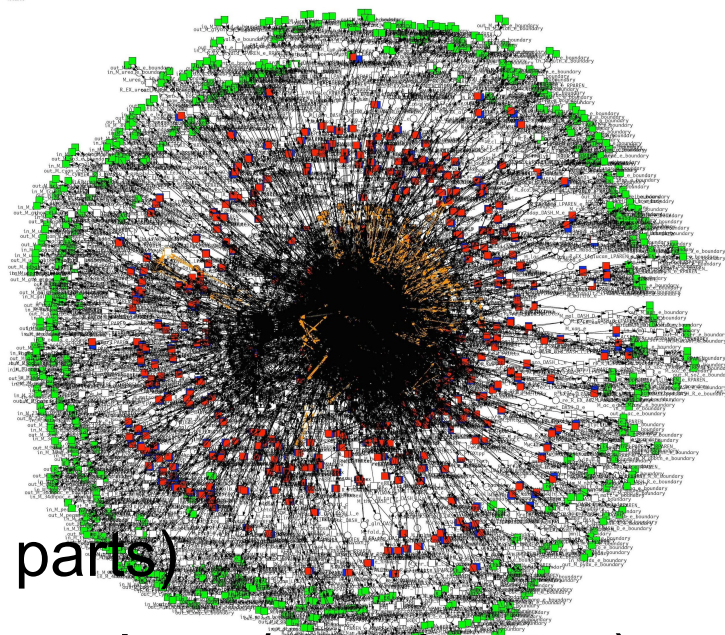


→ cannot perform visual analysis

→ need for automated tools for analysis & correction

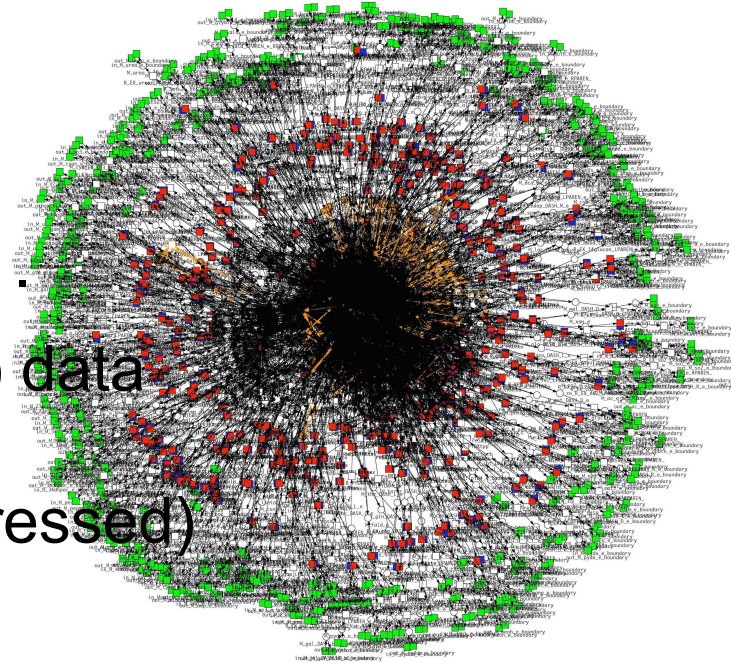
Computational Challenges (2)

- models constructed manually
 - **possibility of 'errors'**
 - typos
 - wrong directions
 - missing information (reactions & metabolites / graph parts)
 - incorrect information (incorrect reactions / graph parts)
 - incorrect composition of parts (reactions) . . .



Computational Challenges (3)

- graph size & structure
→ computational complexity of structural and dynamic analysis, . . .
- large size of secondary (generated) data
→ simulation traces
(30MB uncompressed/12MB compressed)
- design alternatives
→ generation of (very) many models (thousands) . . .



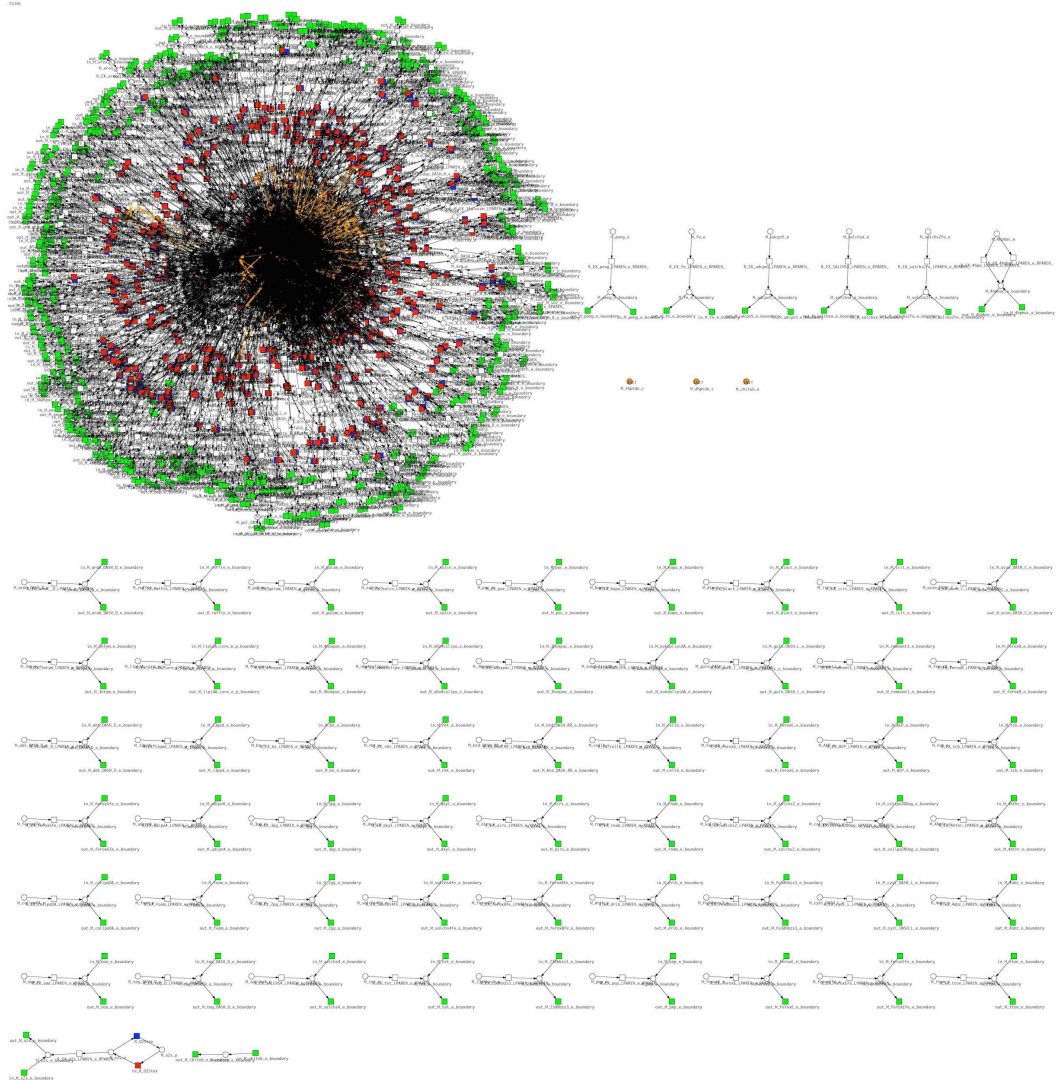
E.coli K-12, MG1655

Whole genome metabolic model

1367 genes
2123 enzymes
2257 metabolites
2645 reactions

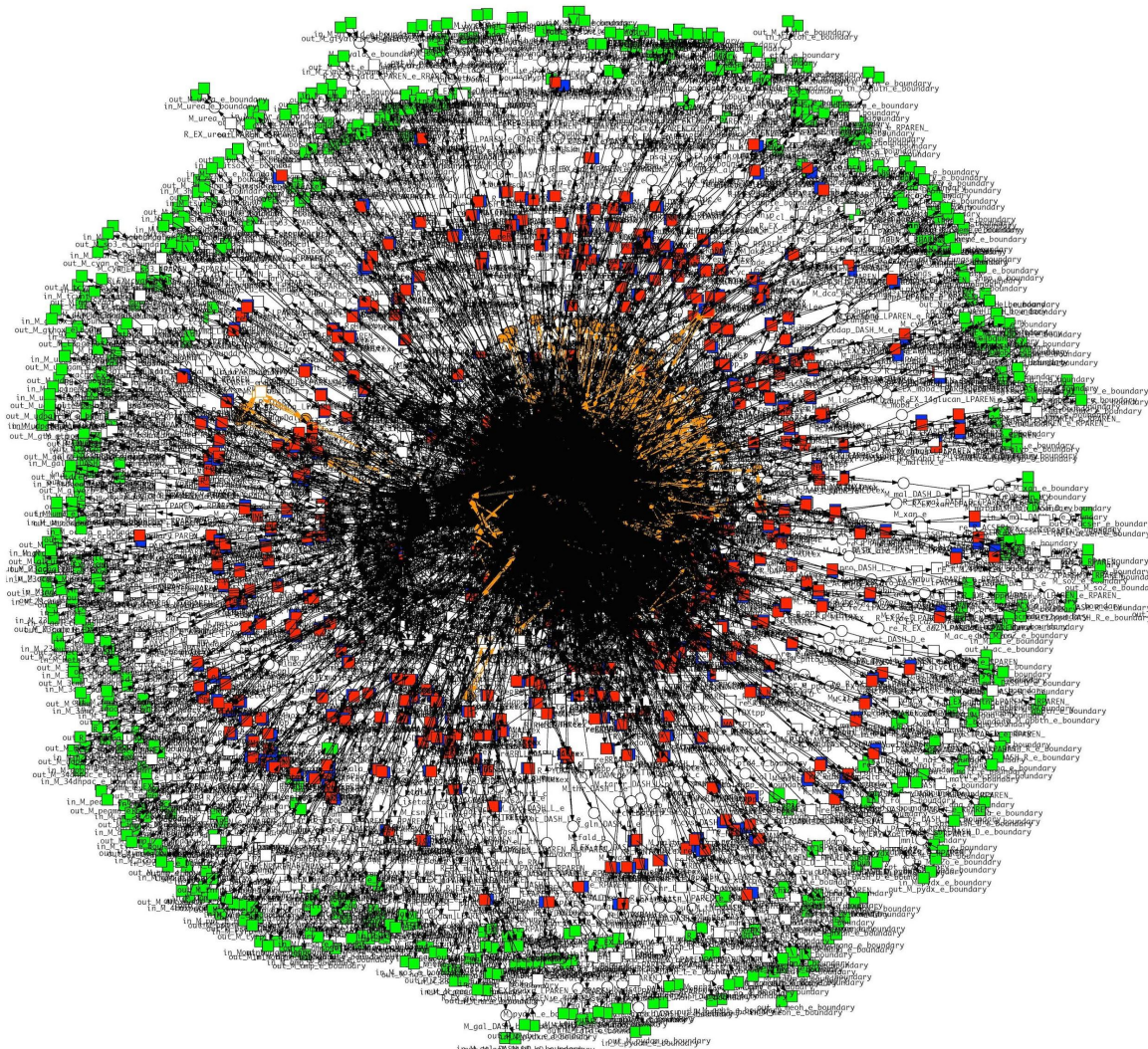
522 spontaneous reactions
11 switched-off reactions
636 reversible reactions
391 boundary conditions

2257 places
4052 transitions



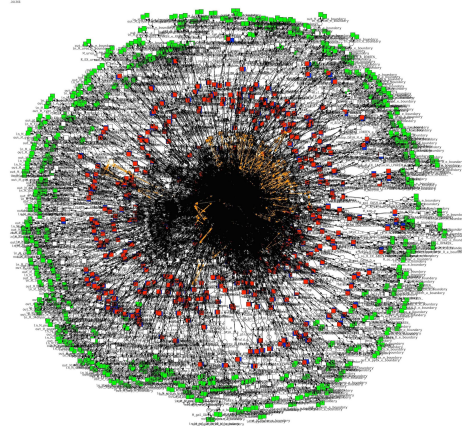
So Big !

We can't
repair this
by hand . . .



Techniques & Tools

- Visualisation & manual editing - **Snoopy**
- Structural analysis
 - **Charlie**
 - ganalysis - **gprolog** (170 predicates / 210 lines)
 - **LoLA** (SAT checker Minisat)
- Automated graph editing
 - 'the protocol' - **gprolog** (2k predicates / 2.3k lines), **LoLA** & **Charlie**
- Simulation
 - **Snoopy** (parameter-free; stochastic, continuous)
 - **Marcie** (parameter-free; stochastic)
- Model checking
 - **MC2**
 - **Marcie**



The Workflow

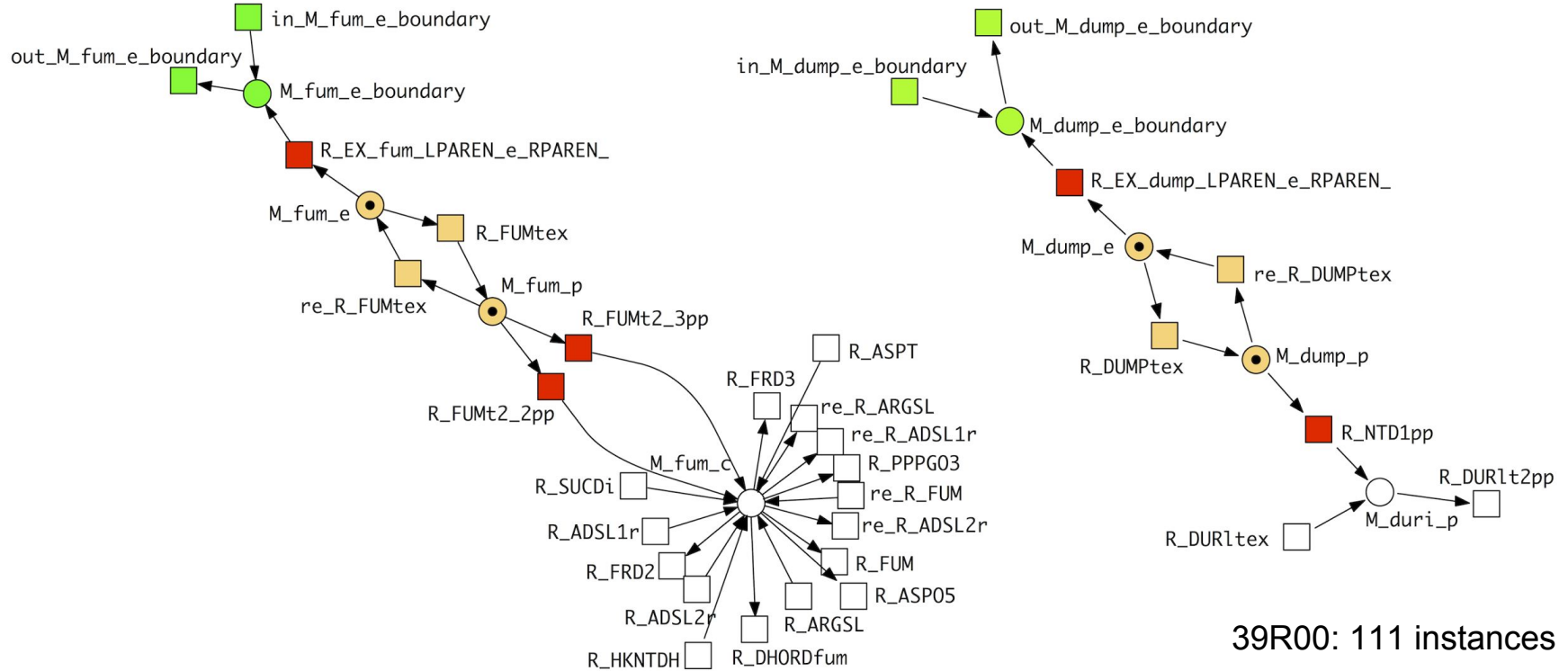
initial model (SBML) → → corrected model

- SBML → Petri net (Snoopy)
 - add boundary reactions (in/out flow) for all boundary conditions
 - reversible reactions → 2*1-way reactions
 - export to graph format (andl)
- Initialise initial model (P-invariants), simulate & analyse
- Automated model correction
- Initialise final model (P-invariants), simulate & analyse
- Compare initial & final models' behaviour

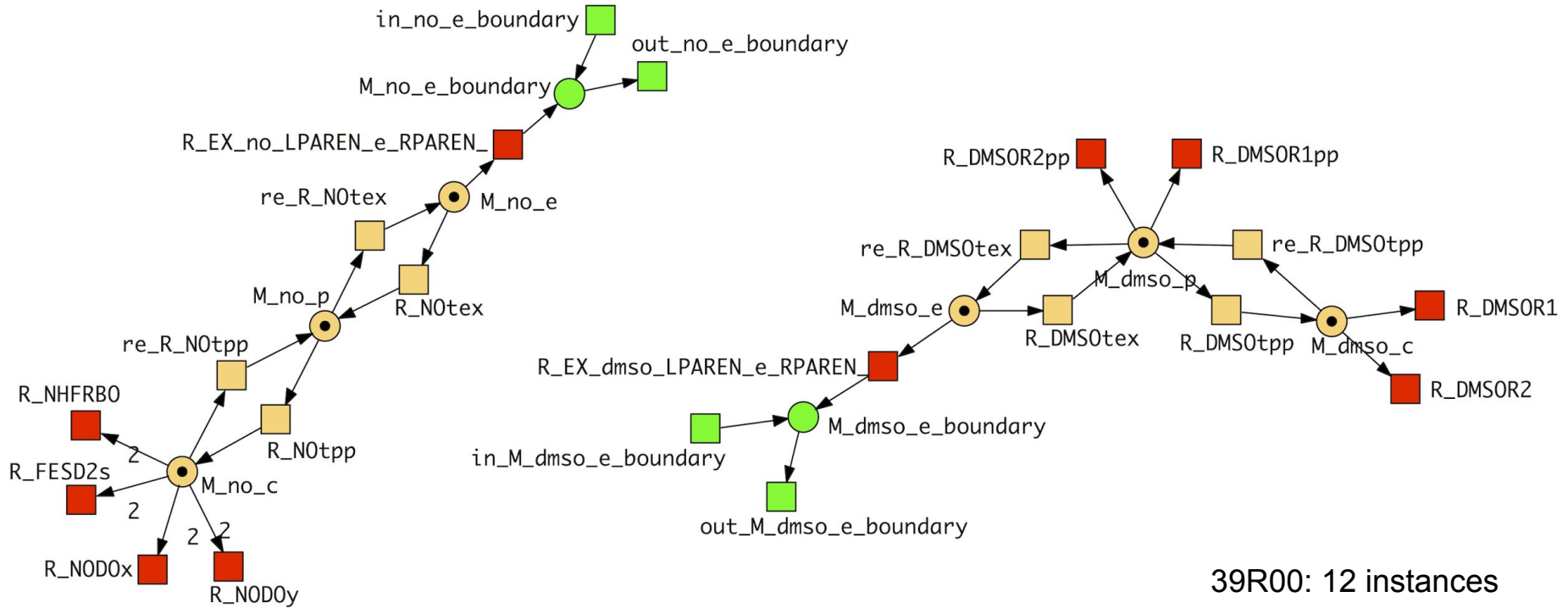
Automated Model Correction

1. Delete isolated nodes (metabolites)
2. Delete minor subnets
3. Repair source/sink places
4. Fix minimal bad siphons by pattern search
5. Fix remaining minimal bad siphons using LoLA+Charlie
6. Check for & delete parallel transitions
7. Reduce (length/complexity of) flow paths
8. Populate the P-invariants in final models
9. Output final model & change log

Some Bad Siphon Patterns (1)

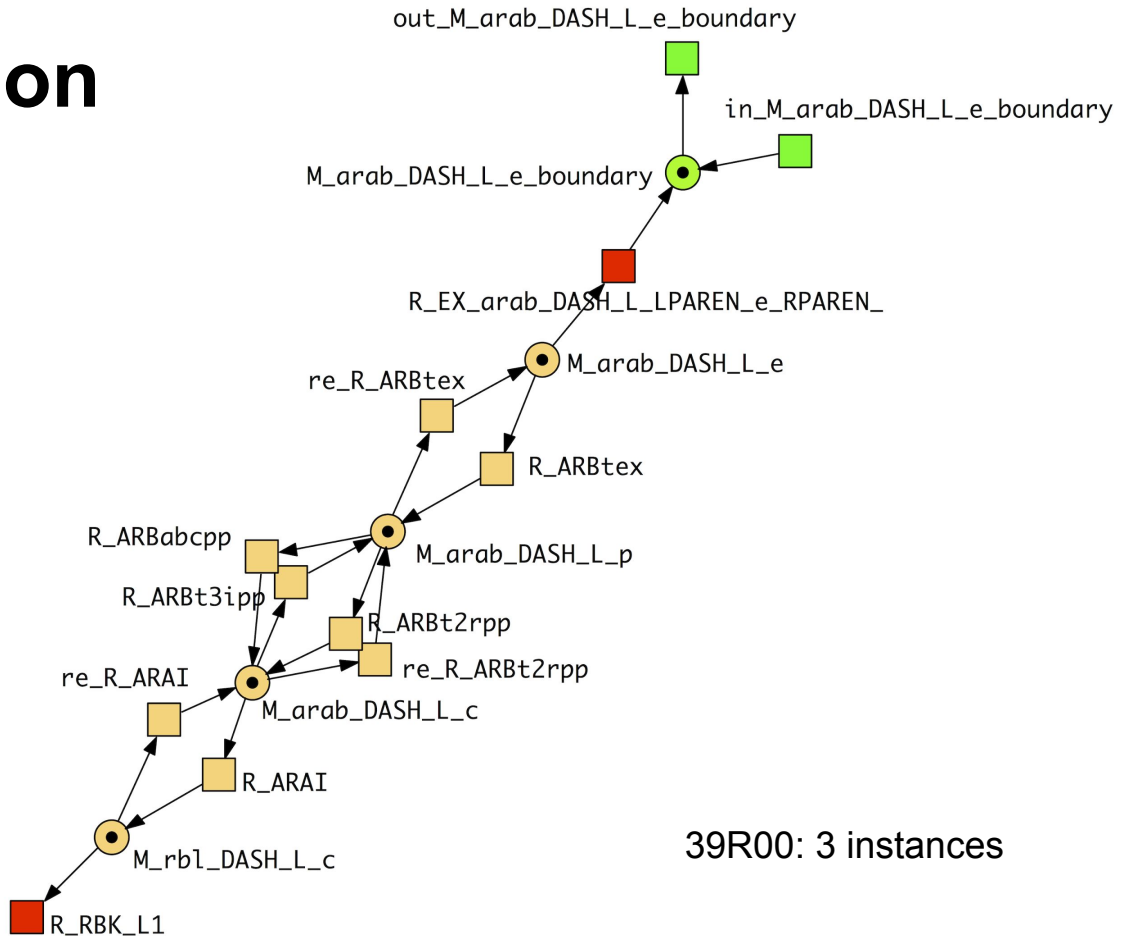


Some Bad Siphon Patterns (2)



39R00: 12 instances

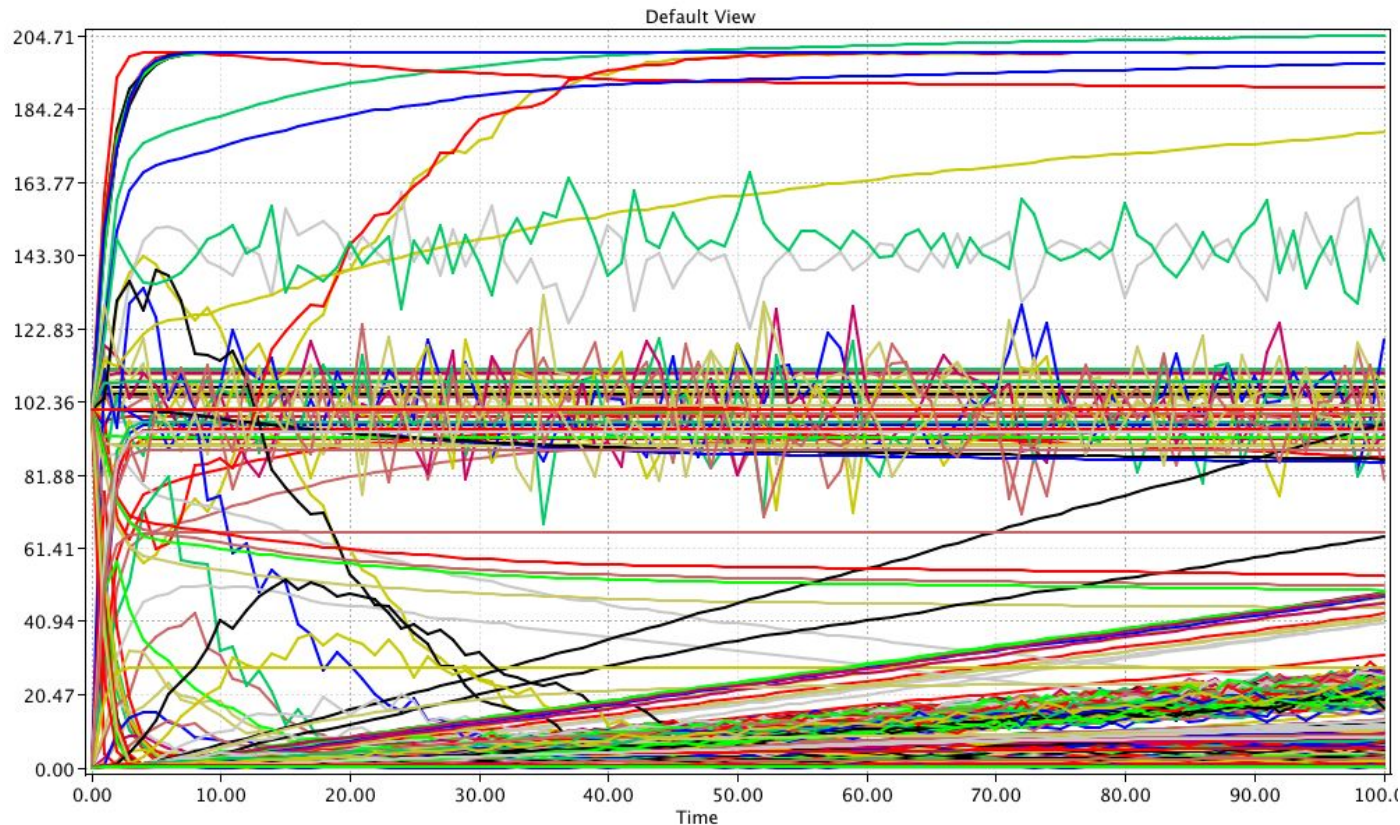
Some Bad Siphon Patterns (3)



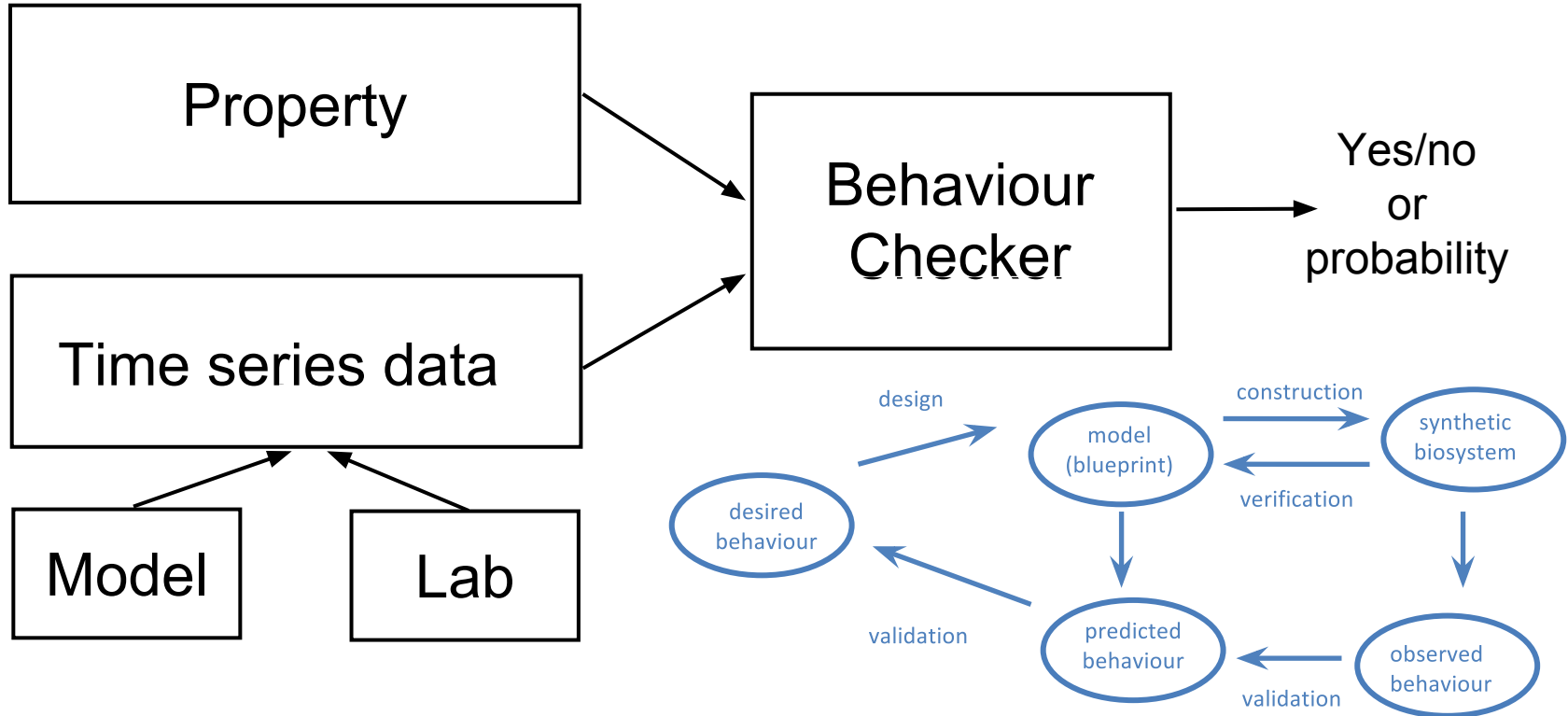
Basic Editing Operations

- Delete metabolite from overall model
- Add/delete metabolite to/from reaction
- Delete reaction
- Reverse reaction (substrates \longleftrightarrow products)
- Add the reverse to a reaction

Time Series for all Metabolites



Simulation-based Model Checking



PLTL properties - Metabolites

$P \geq 1 [G (x=0)]$

% 01_always_steadystate_zero

$P \geq 1 [G (d(x)=0 \wedge x>0)]$

% 02_always_steadystate_above_zero

$P \geq 1 [G (d(x)=0)]$

% 03_always_steadystate_any_value

$P \geq 1 [F (G (x=0 \wedge d(x)=0)) \wedge F (d(x) \neq 0)]$ % 04_changing_and_finally_steadystate_of_zero

$P \geq 1 [F (G (x>0 \wedge d(x)=0)) \wedge F (d(x) \neq 0)]$ % 05_changing_and_finally_steadystate_above_zero

$P \geq 1 [G (d(x)<0)]$

% 07a_decreasing

$P \geq 1 [G (d(x)>0)]$

% 08a_increasing

$P \geq 1 [F(d(x)>0) \wedge (d(x)>0 \vee (G d(x)<0))]$

% 09a_peaks_and_falls

$P \geq 1 [F(d(x)<0) \wedge (d(x)<0 \vee (G d(x)>0))]$

% 10a_falls_and_rises

$P \geq 1 [(F (d(x) \neq 0)) \wedge \neg (F (G (x=0 \wedge d(x)=0)))]$

% 13_activity_and_not_finally_steadystate_of_zero

$P \geq 1 [G (x \leq 0.0001) \wedge \neg G (x=0)]$

% 14a_always_low_concentrations_0.0001

PLTL properties - Reactions

P>=1 [G (x=0)]

P>=1 [F (x>0)]

P>=1 [G (d(x) = 0)]

P>=1 [F (G (x>0))]

P>=1 [F (G (x>0 ^ d(x)=0))]

P>=1 [G (F (x>0))]

P>=1 [F (G (x=0))]

P>=1 [G (d(x)<0)]

P>=1 [G (d(x)>0)]

P>=1 [F(d(x)>0) ^ (d(x)>0 U (G d(x)<0))]

P>=1 [F(d(x)<0) ^ (d(x)<0 U (G d(x)>0))]

P>=1 [G (x<=0.0001) ^ ¬ G (x=0)]

% 01_never_active

% 02_sometime_active

% 04_always_steadystate_active_any_value

% 05a_finally_active

% 05b_finally_active_steadystate

% 05c_always_active_again

% 06_finally_inactive

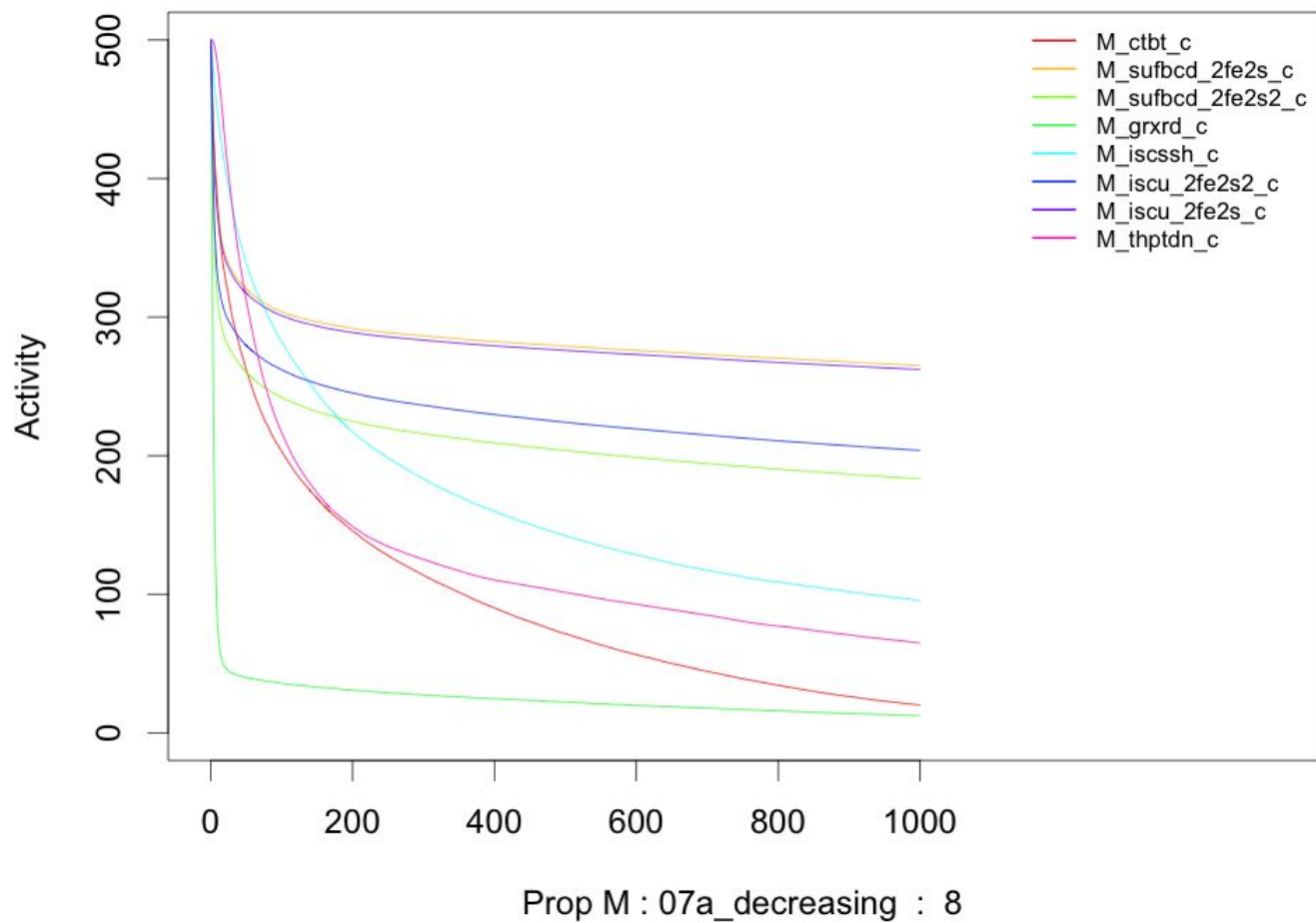
% 07a_always_decreasing_activity

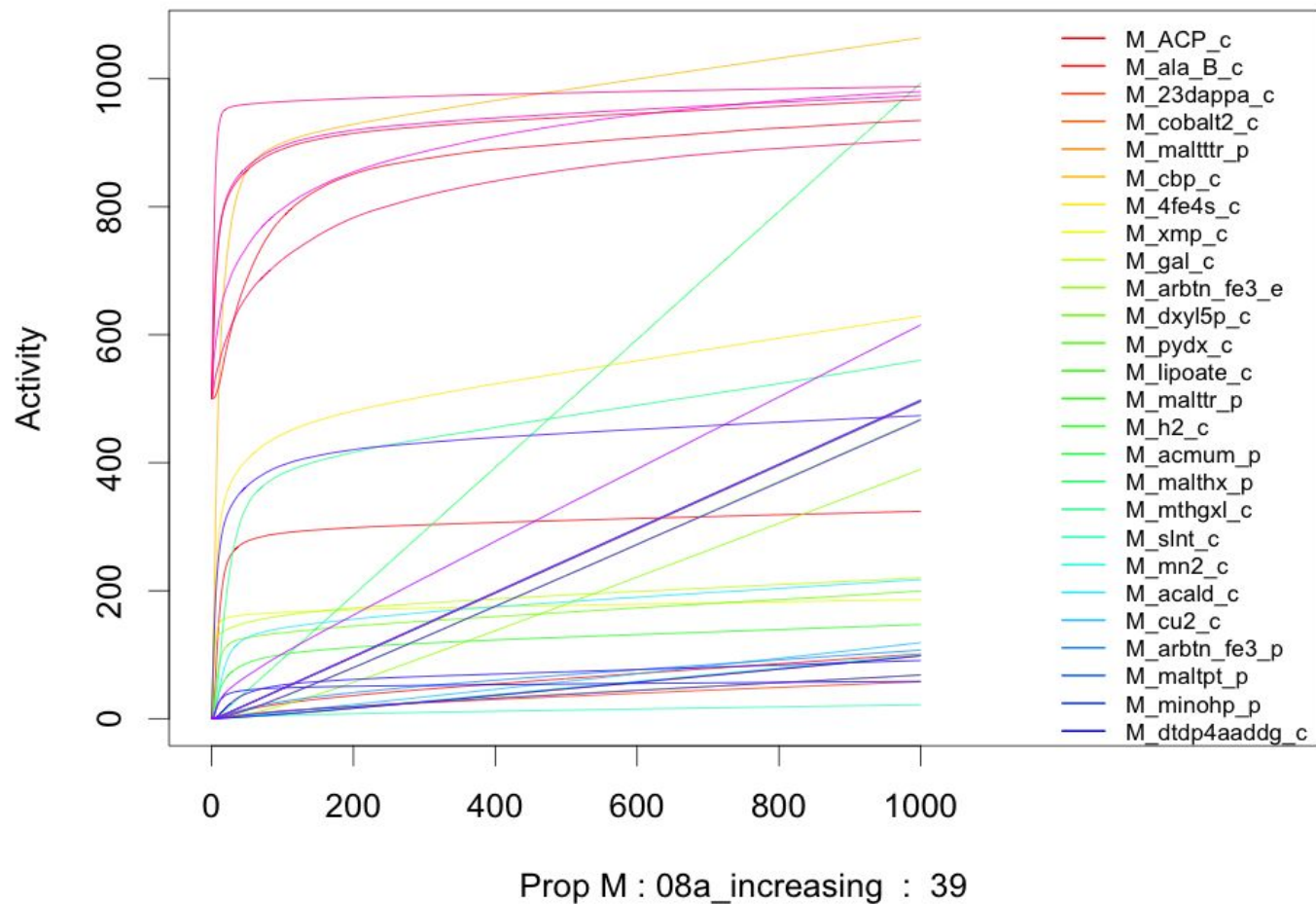
% 08a_always_increasing_activity

% 09a_activity_peaks_and_falls

% 10a_activity_falls_and_rises

% 14a_rare_events_0.0001

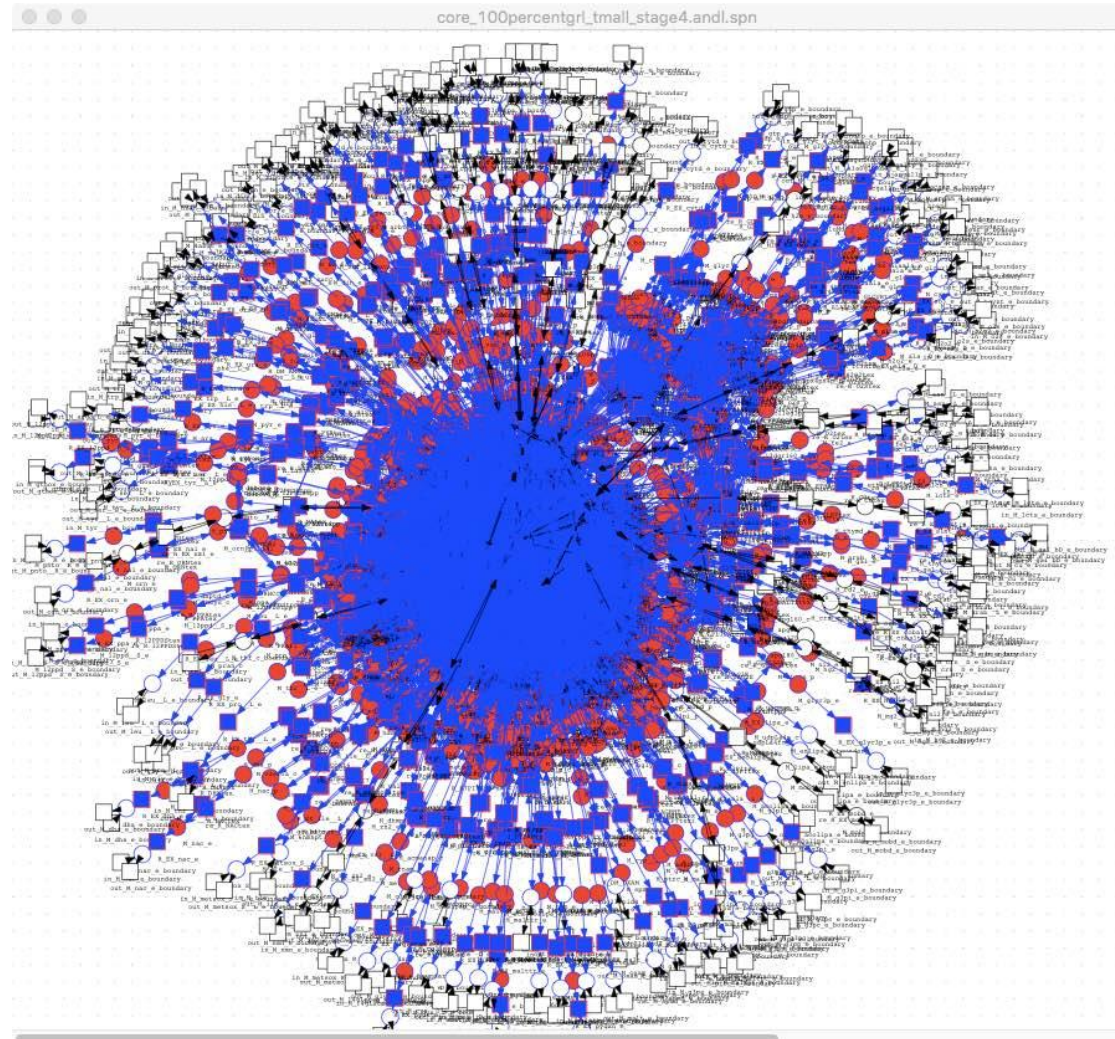




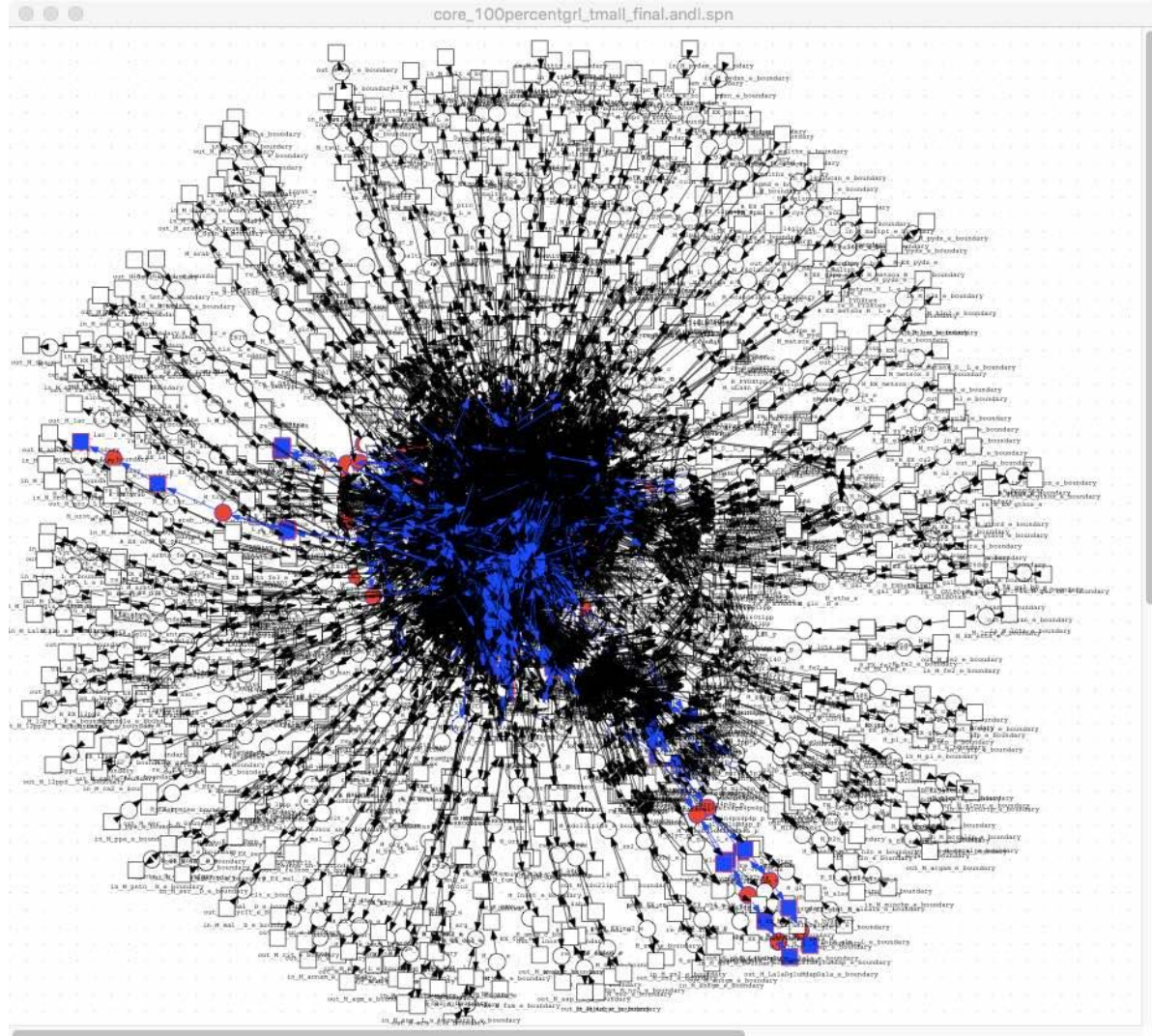
Dead Networks

- All dead metabolites
(M03 - always steady state any value)
& the reactions for which they are substrates/products
- All dead reactions
(R01 - never active)
& their substrates + products

Dead network before repair



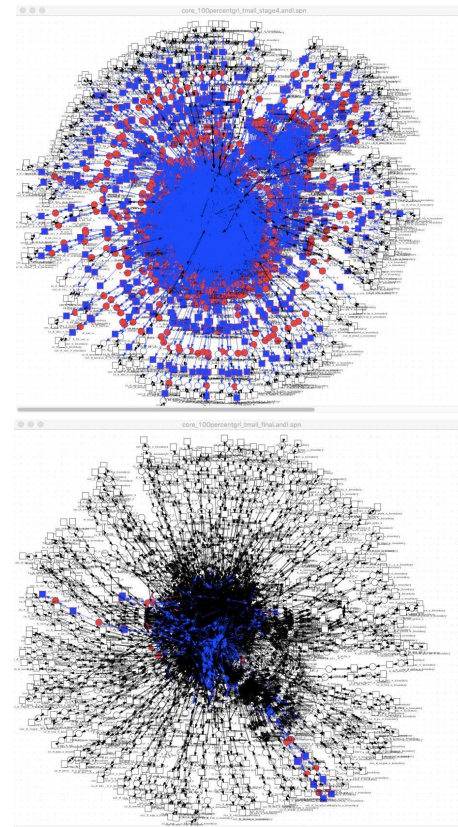
Dead network after repair



E. coli K-12 - Model Repair Summary

	initial model	repaired model
components	66	1
metabolites	2257	2135
reactions	4052	4184
arcs	13,227	13,648
reversible reactions	636	858
boundary conditions	391	323
source & sink places	192	0

bad siphons fixed by LoLA: 28
troublemakers fixed: 177
protocol runtime: 1h



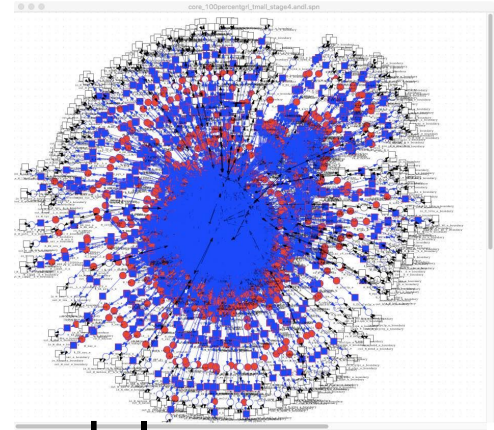
Conclusions

What we achieved so far:

- automated correction protocol for bacterial whole genome metabolic models
- set of analytical tools & techniques
- model database

Side-effects:

- tool improvements
- integration within the synthetic biology theme



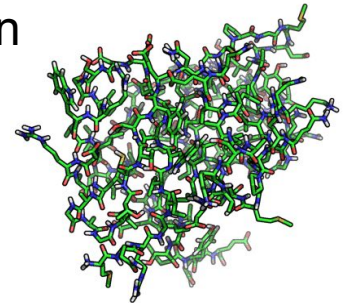
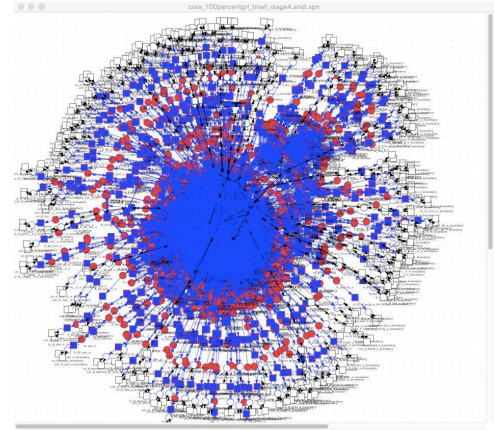
Carrying on

Carrying on



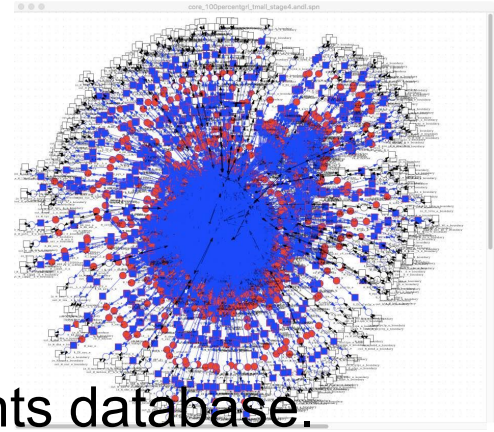
Carrying on

- Improve correction of networks beyond bad siphons (dead nets)
- Gap filling:
finding missing reactions & metabolites due to
 - genes found but reactions missing in the Monk 55 data set
 - genes/reactions not found due to errors in sequencing etc
 - incomplete knowledge of gene-protein-reaction relation
- Extend model to multiscale by including protein structure (with Alessandro Pandini)



The Future !

- Develop method[s] to **optimise design of bacterial strains** using the constructed models & Brunel's model components database.
- **Select appropriate strain & donor alleles/genes** from other strains to optimise
 - target[s] production
 - ease/cost of gene transfer
 - gen[om]e stability
- Identify **genes to modify** to further enhance target achievement



The Team

- David Gilbert
- Monika Heiner
- Bello Suleiman
- Yasoda Jayaweera
- Alessandro Pandini
- Crina Grosan
- Nigel Saunders
- Arshad Khan

Thanks to

CEDPS

- *Supporting MH's visit*
- *Compute power*

BTU Cottbus

- *Christian Rohr*
- *Mostafa Herajy*

Uni Rostock

- *Karsten Wolf (LoLA)*

Questions?



P / T - invariants



- P-invariants:
 - mass conservation
- T-invariants:
 - cyclic behaviour
 - steady state

