Modelling tools for Bio-PEPA

Stephen Gilmore

Centre for Systems Biology at Edinburgh
The University of Edinburgh
Edinburgh, EH9 3JH, U.K.

Joint work with
Federica Ciocchetta, Allan Clark, Adam Duguid, Vashti Galpin, Maria Luisa Guerriero, Jane Hillston and Laurence Loewe

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

- Properties
  - Formal
  - High-level
  - Concise
  - Readable
  - Static analysis
  - Multiple analysis vectors

Ciocchetta, F., and J. Hillston.
Bio-PEPA: A framework for the modelling and analysis of biological systems.
Theoretical Computer Science.
Volume 410, Issues 33-34, 21 August 2009, Pages 3065–3084.
Concurrent Systems Biology: To Nadia Busi (1968–2007)

Discrete and stochastic or continuous and deterministic?

Outline

1. The Bio-PEPA language
2. Bio-PEPA Software Tools
3. Analysis based on ODEs
4. Analysis based on CTMCs
5. Examples: Two Genetic Networks
   - The Network With Protein Degradation ($\mathcal{M}_1$)
   - The Network Without Protein Degradation ($\mathcal{M}_2$)
6. Larger examples

Outline

- Consider the simple Enzyme-Substrate reaction involving an enzyme E, a substrate S, a compound $E{:}S$ and a product P.

$$\boxed{E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E{:}S \overset{k_2}{\rightarrow} E + P}$$

The kinetic functions

r_1	$k_1 \times E \times S$
r_{-1}	$k_{-1} \times E{:}S$
r_2	$k_2 \times E{:}S$

The Bio-PEPA model

E	$r_1 \downarrow$	$+$	$r_{-1} \uparrow$	$+$	$r_2 \uparrow$
S	$r_1 \downarrow$	$+$	$r_{-1} \uparrow$		
$E{:}S$	$r_1 \uparrow$	$+$	$r_{-1} \downarrow$	$+$	$r_2 \downarrow$
P					$r_2 \uparrow$

The Bio-PEPA model

E	$r_1 \downarrow$	$+$	$r_{-1} \uparrow$	$+$	$r_2 \uparrow$
S	$r_1 \downarrow$	$+$	$r_{-1} \uparrow$		
$E{:}S$	$r_1 \uparrow$	$+$	$r_{-1} \downarrow$	$+$	$r_2 \downarrow$
P					$r_2 \uparrow$

The differential equations

dE/dt	$-r_1$	$+$	r_{-1}	$+$	r_2
dS/dt	$-r_1$	$+$	r_{-1}		
$dE{:}S/dt$	r_1	$-$	r_{-1}	$-$	r_2
dP/dt					r_2

The Jacobian

The differential equations

dE/dt	$-k_1 \times E \times S + k_{-1} \times E{:}S + k_2 \times E{:}S$
dS/dt	$-k_1 \times E \times S + k_{-1} \times E{:}S$
$dE{:}S/dt$	$k_1 \times E \times S - k_{-1} \times E{:}S - k_2 \times E{:}S$
dP/dt	$k_2 \times E{:}S$

The Jacobian

	E	S	$E{:}S$	P
E	$\partial f_E/\partial E$	$\partial f_E/\partial S$	$\partial f_E/\partial E{:}S$	$\partial f_E/\partial P$
S	$\partial f_S/\partial E$	$\partial f_S/\partial S$	$\partial f_S/\partial E{:}S$	$\partial f_S/\partial P$
$E{:}S$	$\partial f_{E:S}/\partial E$	$\partial f_{E:S}/\partial S$	$\partial f_{E:S}/\partial E{:}S$	$\partial f_{E:S}/\partial P$
P	$\partial f_P/\partial E$	$\partial f_P/\partial S$	$\partial f_P/\partial E{:}S$	$\partial f_P/\partial P$

The Jacobian

The differential equations

dE/dt	$-k_1 \times E \times S + k_{-1} \times E{:}S + k_2 \times E{:}S$
dS/dt	$-k_1 \times E \times S + k_{-1} \times E{:}S$
$dE{:}S/dt$	$k_1 \times E \times S - k_{-1} \times E{:}S - k_2 \times E{:}S$
dP/dt	$k_2 \times E{:}S$

The Jacobian

	E	S	$E{:}S$	P
E	$-k_1 \times S$	$-k_1 \times E$	$k_{-1} + k_2$	
S	$-k_1 \times S$	$-k_1 \times E$	k_{-1}	
$E{:}S$	$k_1 \times S$	$k_1 \times E$	$-k_{-1} - k_2$	
P			k_2	

Using the Jacobian

- ODE solvers generally use *finite differences* to approximate the Jacobian matrix if it is not supplied, but an implementation of the analytically derived Jacobian can improve the speed, accuracy and reliability of the program.

Using the Jacobian

- ODE solvers generally use *finite differences* to approximate the Jacobian matrix if it is not supplied, but an implementation of the analytically derived Jacobian can improve the speed, accuracy and reliability of the program.
- The Jacobian (and Hessian and higher derivatives) are computed automatically from the differential equations using *symbolic differentiation*.

- ODE solvers generally use *finite differences* to approximate the Jacobian matrix if it is not supplied, but an implementation of the analytically derived Jacobian can improve the speed, accuracy and reliability of the program.
- The Jacobian (and Hessian and higher derivatives) are computed automatically from the differential equations using *symbolic differentiation*.
- Programs that compute *bifurcations* will use the Hessian and higher derivatives.

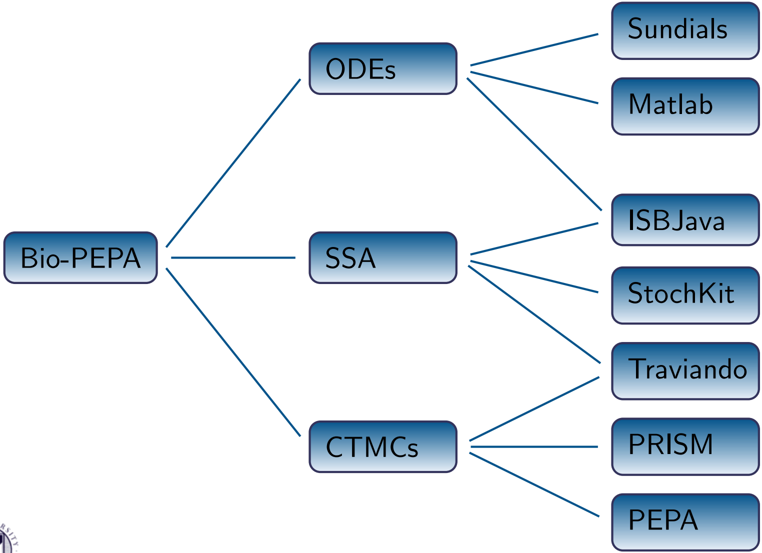
Outline

Requirements

1. Facilitate running several different types of quantitative analysis on a single Bio-PEPA model.
2. Facilitate combining the results of several runs of one particular type of analysis of a Bio-PEPA model.
3. Require as little explicit programming as possible from users.
4. Allow users to choose the parameters of interest for closer investigation.
5. Build on other mature simulators and numerical libraries where possible to avoid re-implementing existing functionality.
6. Users should be involved in deciding which features are important through suggesting enhancements to current versions.

Bio-PEPA Analysis

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH



Bio-PEPA Eclipse Plug-in

- A complete environment for working with Bio-PEPA models.
- Eclipse front-end and a separate back-end library.

User Interface	Core
editor for the Bio-PEPA language	parser for the Bio-PEPA language
problems view	static analysis
outline view for the reaction-centric view	ISBJava time series analysis (ODE, SSA)
graphing support via common plugin	export facility (SBML; PRISM)

Availability

- The Bio-PEPA tools are freely available for download from `www.biopepa.org`.

Adam Duguid, Stephen Gilmore, Maria Luisa Guerriero, Jane Hillston and Laurence Loewe.

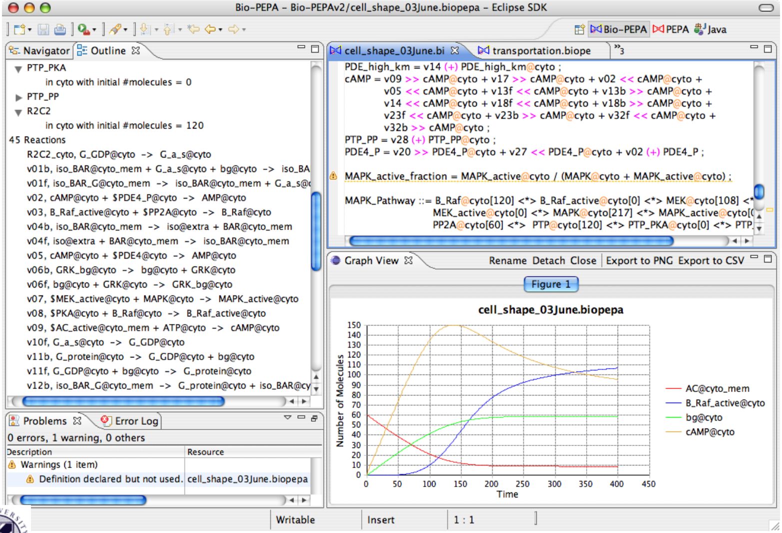
Design and Development of Software Tools for Bio-PEPA.

Winter Simulation Conference.

Austin, Texas. December 2009.

Software demo:
Bio-PEPA Eclipse Plug-in

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH



Outline

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

- Several different differential equation solvers exist.
  - SUNDIALS — ODE integrators in C
  - Matlab — numerical computing platform
  - MatCont — Matlab toolbox for continuation analysis
  - AUTO — C and Fortran package for numerical continuation
- Different formats and languages for problem description.

Using VFGEN

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

- VFGEN is a vector field file generator for differential equation solvers and other computational tools.
- VFGEN lets you define your vector field once (using XML), and export the vector field in several formats.
- VFGEN uses a C++ symbolic algebra library (GiNaC) to generate Jacobians and higher derivatives automatically.

Using VFGEN

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

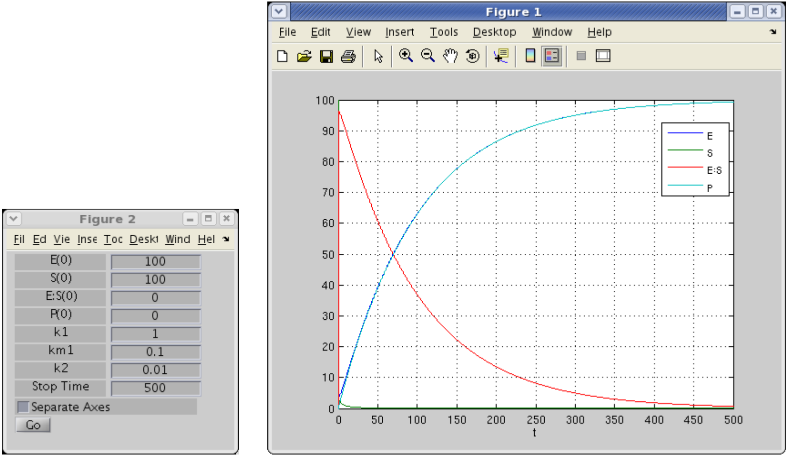
- VFGEN is a vector field file generator for differential equation solvers and other computational tools.
- VFGEN lets you define your vector field once (using XML), and export the vector field in several formats.
- VFGEN uses a C++ symbolic algebra library (GiNaC) to generate Jacobians and higher derivatives automatically.

Warren Weckesser.
VFGEN: A Code Generation Tool.
Journal of Numerical Analysis, Industrial and Applied Mathematics, Volume 3(1-2):151–165, 2008.

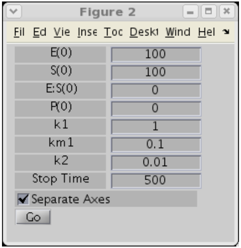
VFgen representation of the Enzyme-Substrate model

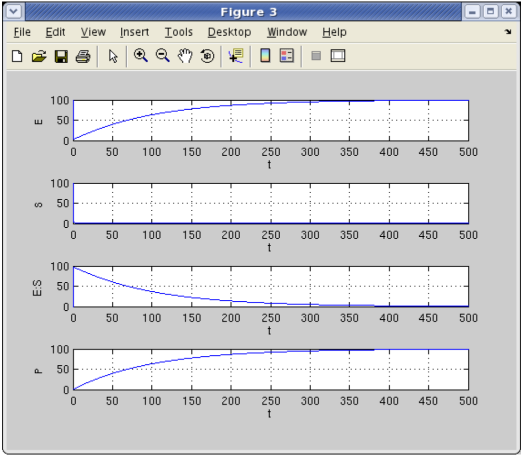

```xml
<?xml version="1.0"?>
<VectorField Name="mm001">

<Parameter Name="k1_" Description="k1" Latex="k_1" DefaultValue="1"/>
<Parameter Name="km1_" Description="km1" Latex="k_{-1}" DefaultValue="0.1"/>
<Parameter Name="k2_" Description="k2" Latex="k_2" DefaultValue="0.01"/>

<Expression Name="r1_" Description="r1" Latex="r_1" Formula=" k1_ * E_ * S_ "/>
<Expression Name="rm1_" Description="rm1" Latex="r_{-1}" Formula=" km1_ * E_colon_S_ "/>
<Expression Name="r2_" Description="r2" Latex="r_2" Formula=" k2_ * E_colon_S_ "/>

<StateVariable Name="E_" Description="E" Latex="E" DefaultInitialCondition="100"
    Formula=" - r1_ + rm1_ + r2_"/>

<StateVariable Name="S_" Description="S" Latex="S" DefaultInitialCondition="100"
    Formula=" - r1_ + rm1_"/>

<StateVariable Name="E_colon_S_" Description="E:S" Latex="\hbox{\textit{E:S}}"
    DefaultInitialCondition="0" Formula="r1_ - rm1_ - r2_"/>

<StateVariable Name="P_" Description="P" Latex="P" DefaultInitialCondition="0"
    Formula="r2_"/>
</VectorField>
```

Analysing Bio-PEPA models with Matlab



Analysing Bio-PEPA models with Matlab

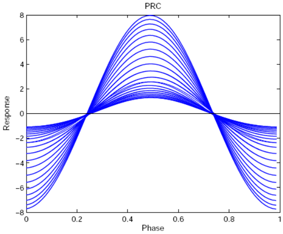


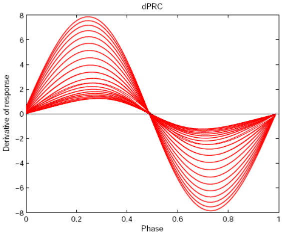


CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

- We have some support for more general analysis of ODE models generated from Bio-PEPA descriptions.
- We can now perform bifurcation analysis and continuation analysis on Bio-PEPA models.
  - Useful for studying systems which oscillate.
- Can compute *phase response curves*.

Phase response curve





A. Dhooge, W. Govaerts, Yu.A. Kuznetsov, W. Mestrom, A.M. Riet and B. Sautois

MATCONT and CL_MATCONT: Continuation toolboxes in Matlab

December 2006.

Outline

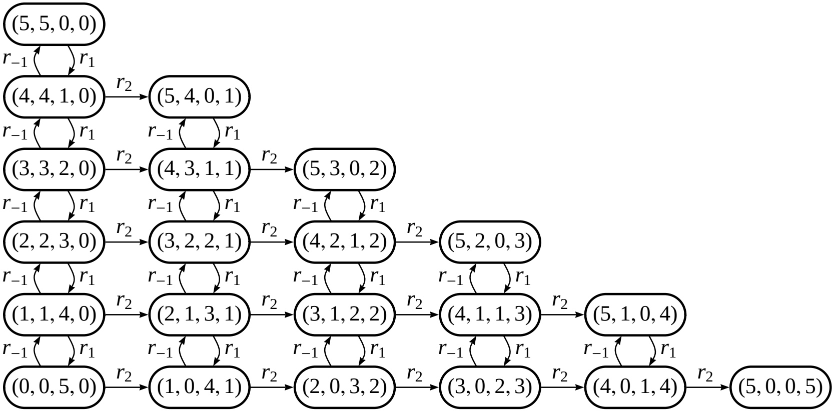
- Consider again the simple Enzyme-Substrate reaction involving an enzyme E, a substrate S, a compound $E{:}S$ and a product P.

$$\boxed{E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E{:}S \overset{k_2}{\rightarrow} E + P}$$

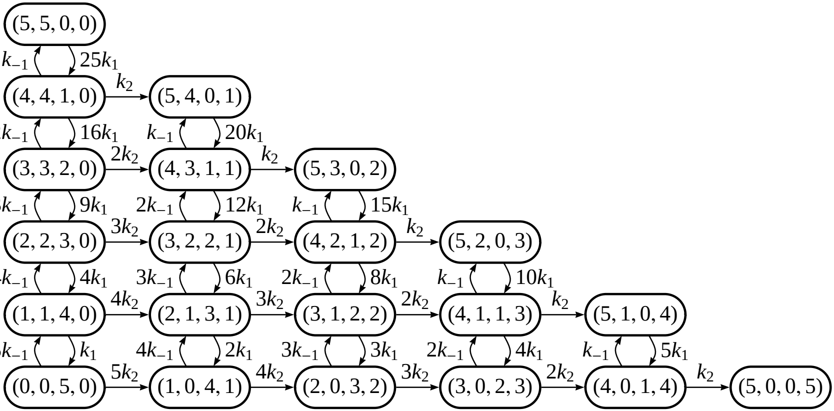
- Suppose that we could initiate this system with only 5 molecules of E, 5 molecules of S, no compound and no product.
- With only 4 species and 3 reaction channels the system has a small reachable state-space.

Discrete state-space of the Enzyme-Substrate example

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH



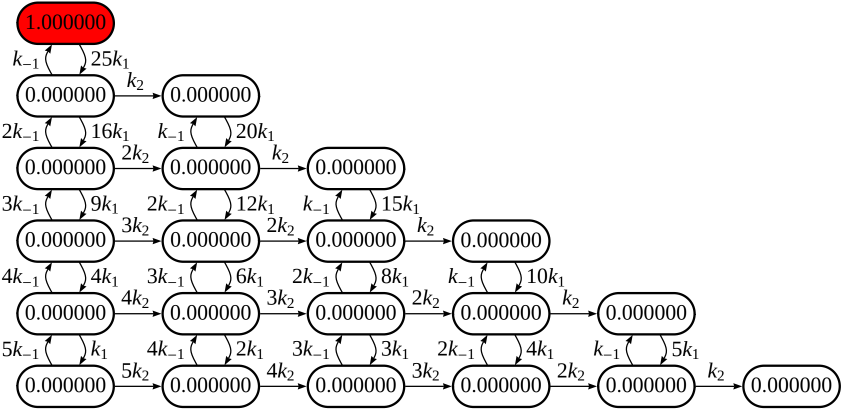
Markov chain of the Enzyme-Substrate example



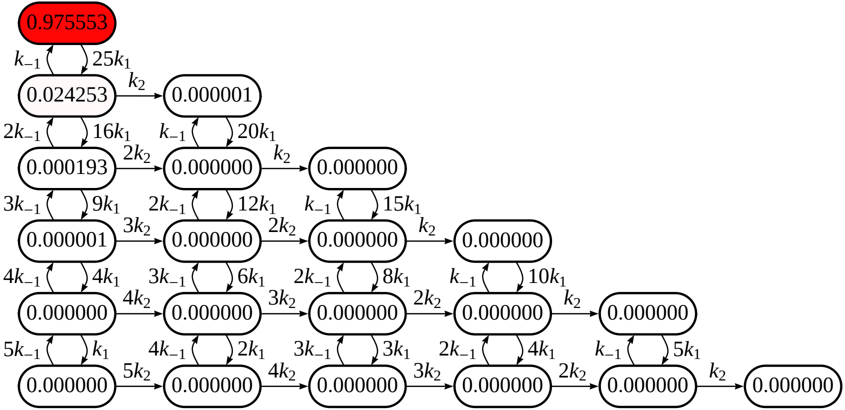
- If we know the initial molecule counts and the values of the rate constants $k_1 = 1.0$, $k_{-1} = 20.0$ and $k_2 = 0.05$ we can compute the probability of being in each state of the state-space at all future time points.
- At time $t = 0$ we have $\Pr(5, 5, 0, 0) = 1$.

Transient probability, $t = 0$

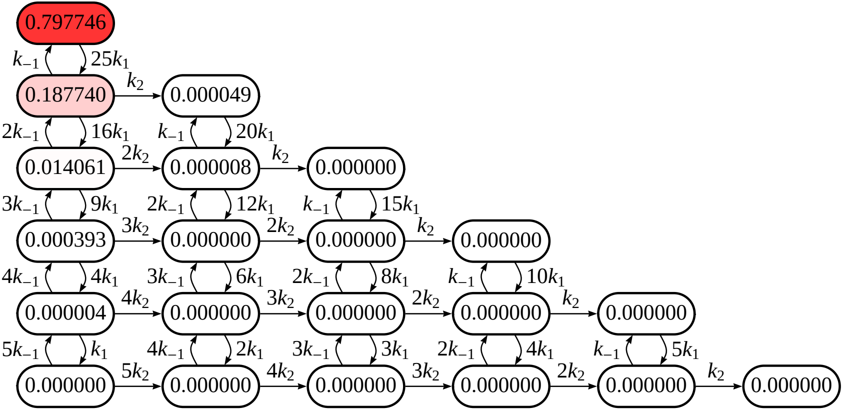
CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH



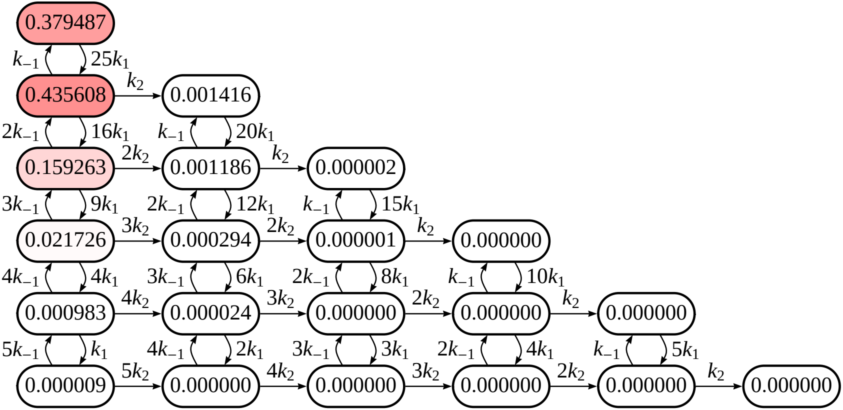
Transient probability, $t = 0.001$



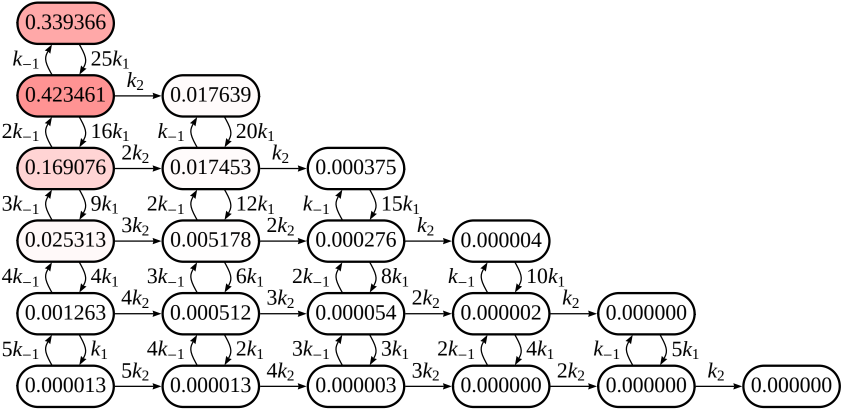
CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH



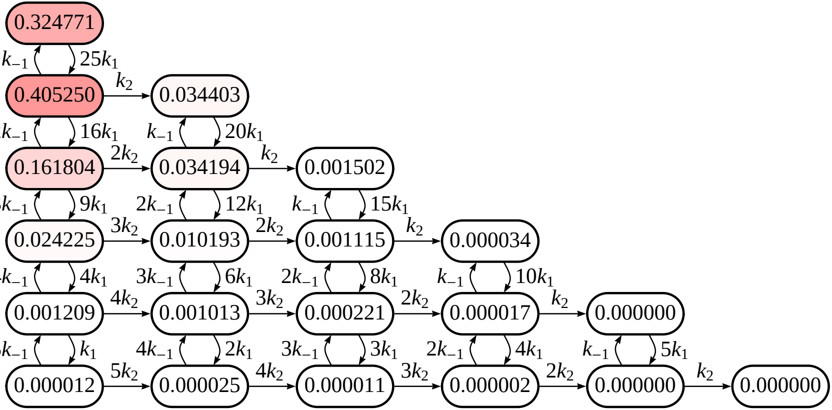
Transient probability, $t = 0.1$



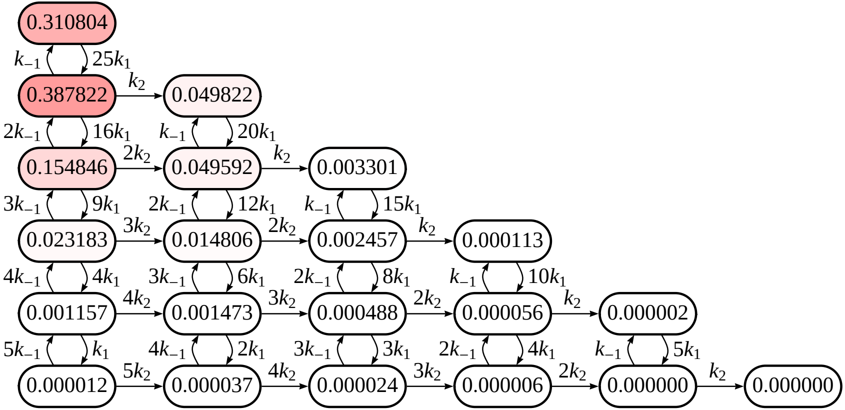
Transient probability, $t = 1$



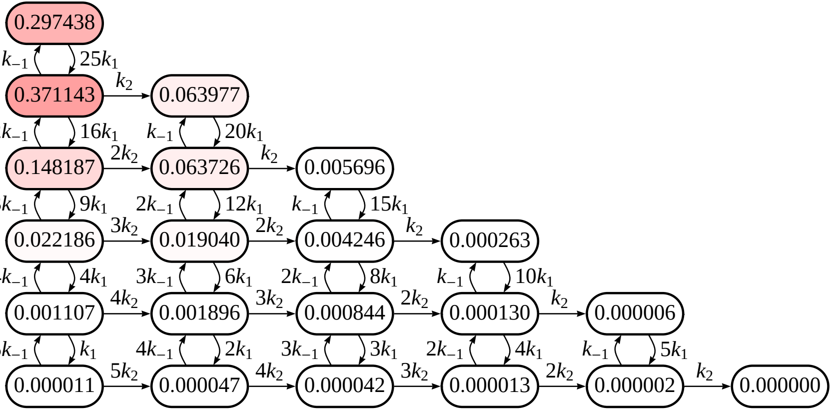
Transient probability, $t = 2$



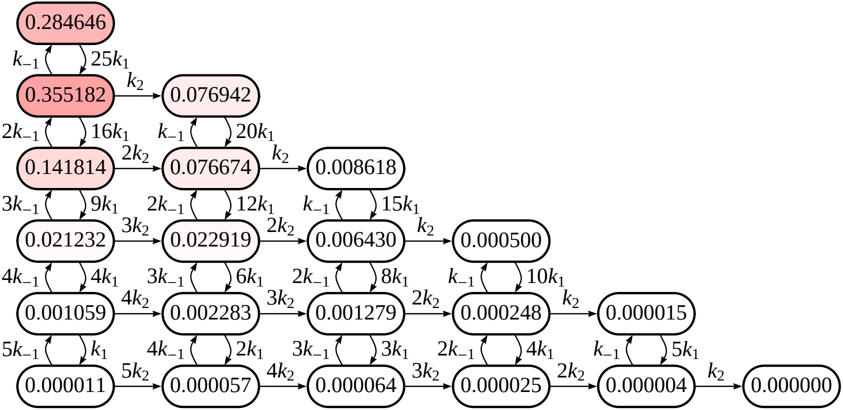
Transient probability, $t = 3$



Transient probability, $t = 4$

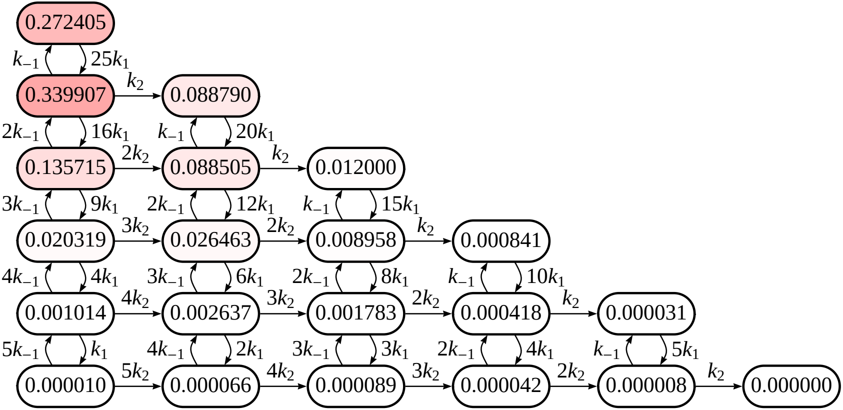


Transient probability, $t = 5$

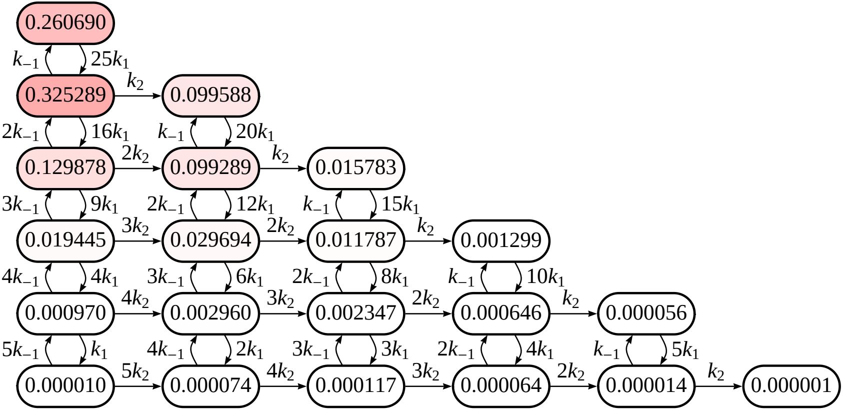


Transient probability, $t = 6$

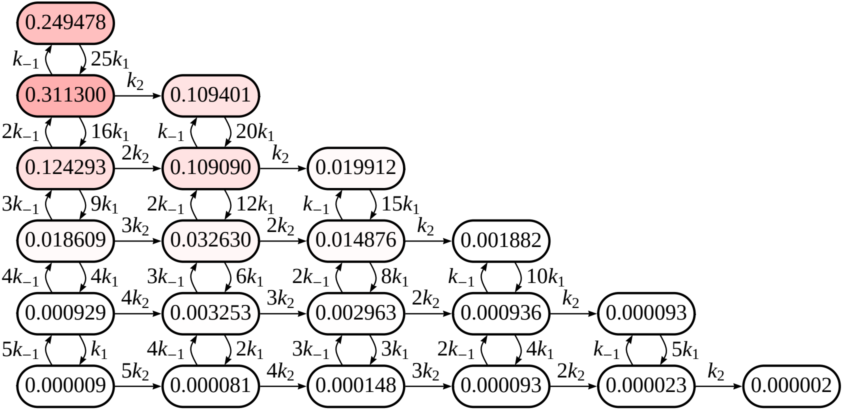
CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH



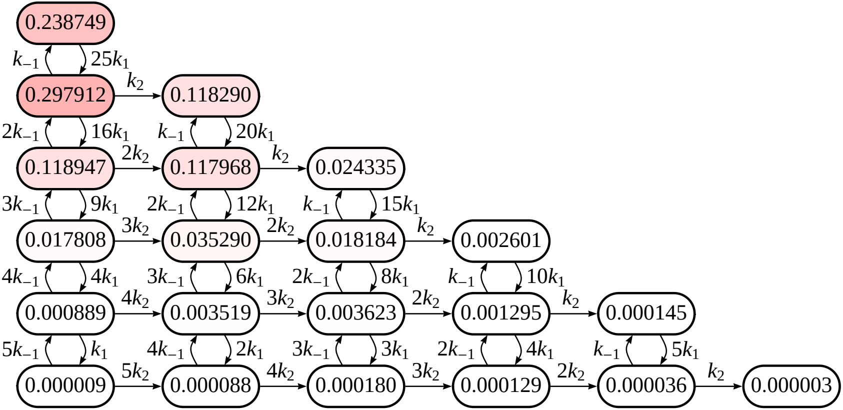
Transient probability, $t = 7$

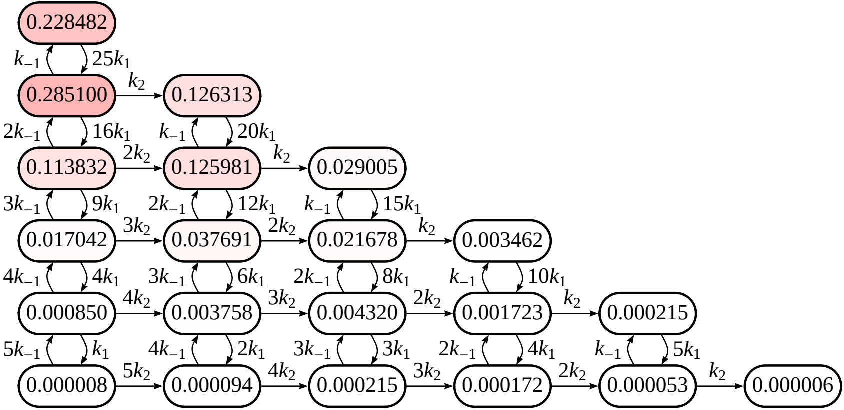


Transient probability, $t = 8$

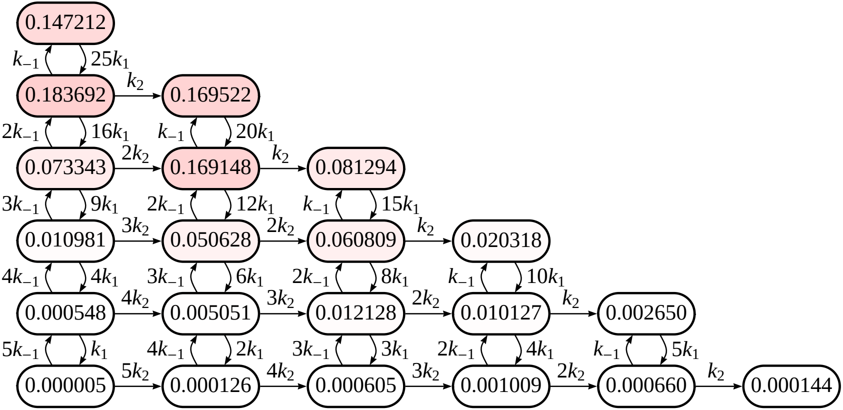


Transient probability, $t = 9$

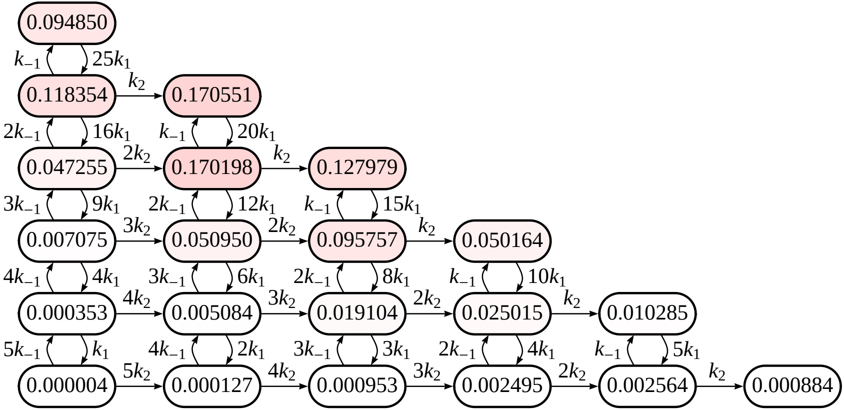



0.228482
k_{-1}
$25k_1$
0.285100
k_2
0.126313
$2k_{-1}$
$16k_1$
k_{-1}
$20k_1$
0.113832
$2k_2$
0.125981
k_2
0.029005
$3k_{-1}$
$9k_1$
$2k_{-1}$
$12k_1$
k_{-1}
$15k_1$
0.017042
$3k_2$
0.037691
$2k_2$
0.021678
k_2
0.003462
$4k_{-1}$
$4k_1$
$3k_{-1}$
$6k_1$
$2k_{-1}$
$8k_1$
k_{-1}
$10k_1$
0.000850
$4k_2$
0.003758
$3k_2$
0.004320
$2k_2$
0.001723
k_2
0.000215
$5k_{-1}$
k_1
$4k_{-1}$
$2k_1$
$3k_{-1}$
$3k_1$
$2k_{-1}$
$4k_1$
k_{-1}
$5k_1$
0.000008
$5k_2$
0.000094
$4k_2$
0.000215
$3k_2$
0.000172
$2k_2$
0.000053
k_2
0.000006

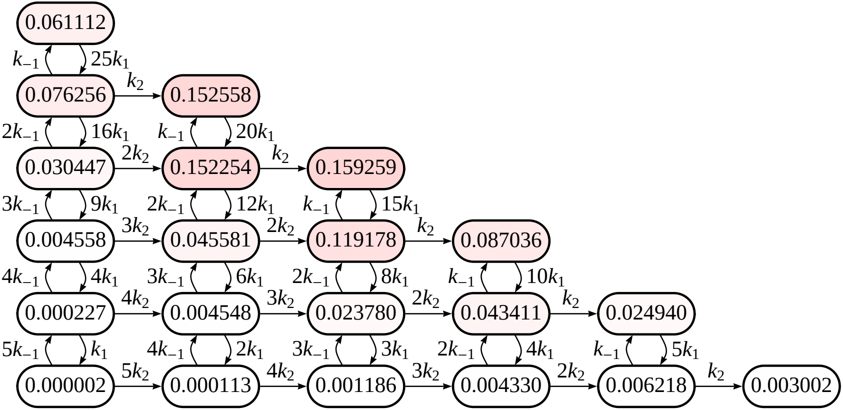
Transient probability, $t = 20$



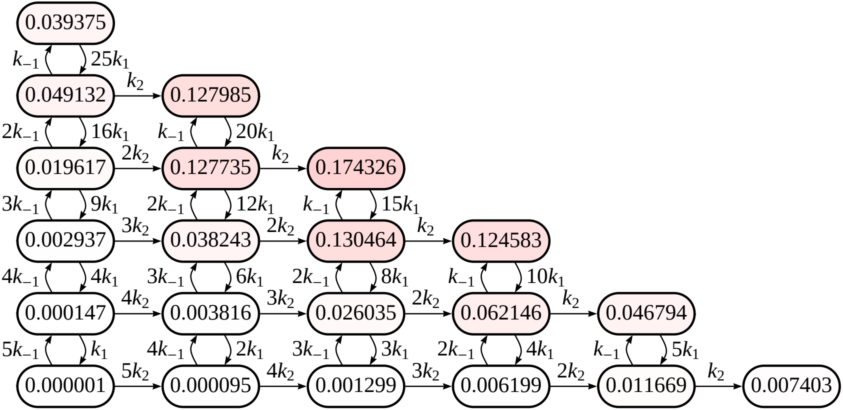
Transient probability, $t = 30$



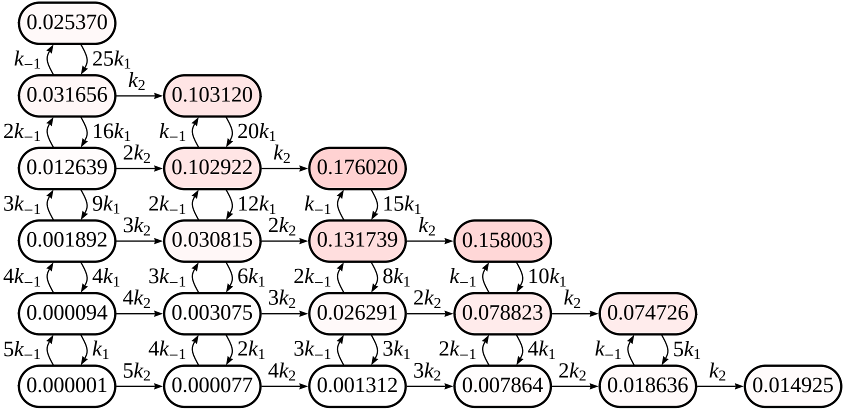
Transient probability, $t = 40$



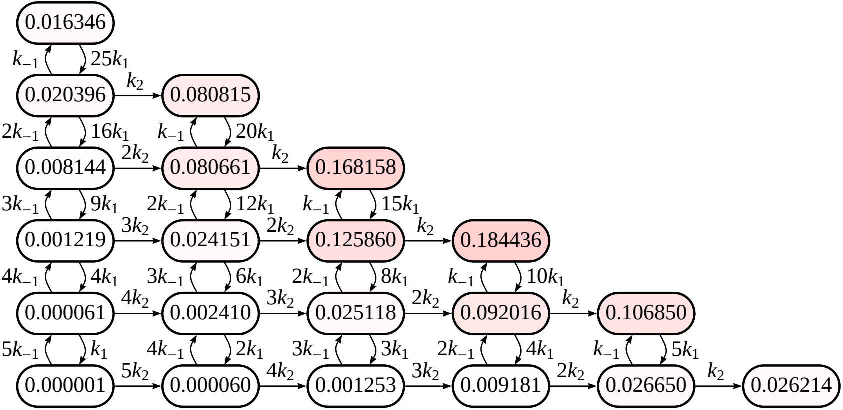
Transient probability, $t = 50$



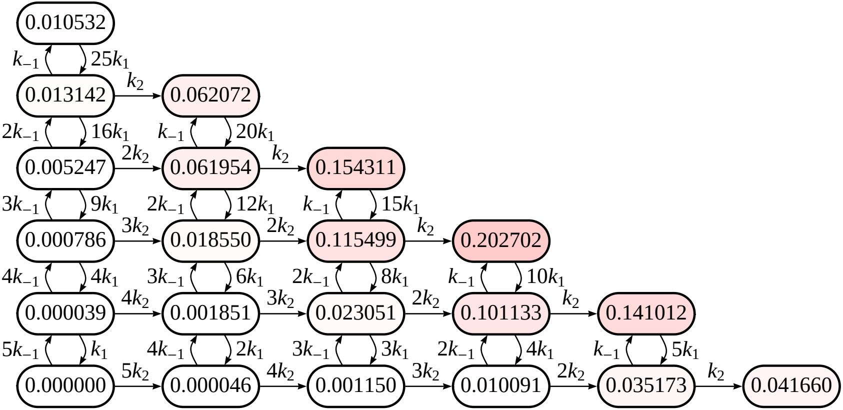
Transient probability, $t = 60$



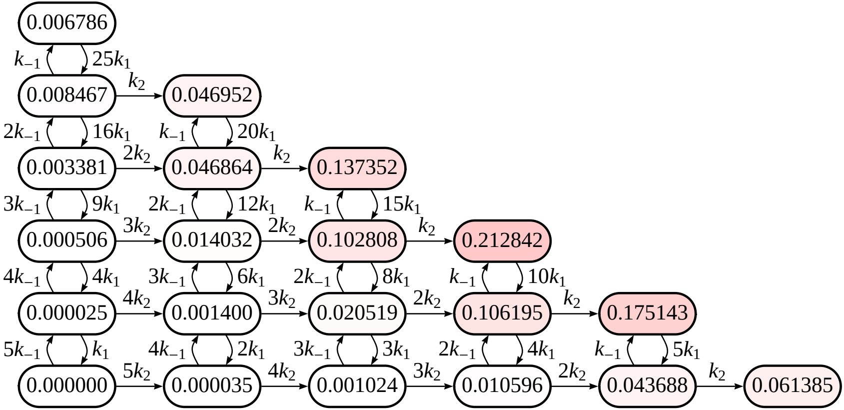
Transient probability, $t = 70$



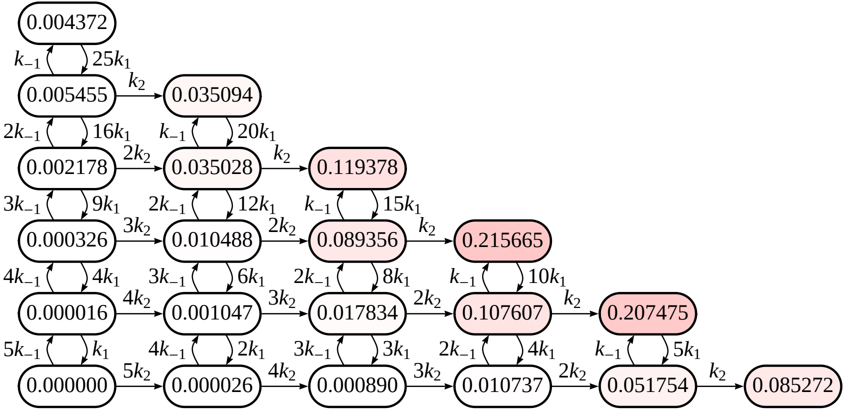
Transient probability, $t = 80$



CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

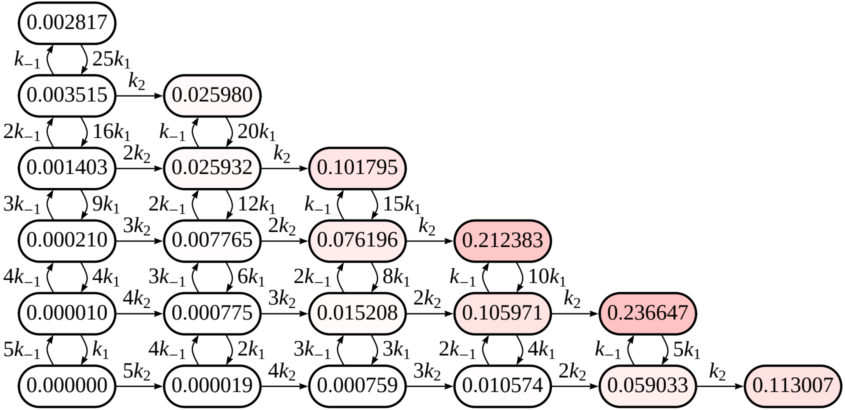


Transient probability, $t = 100$

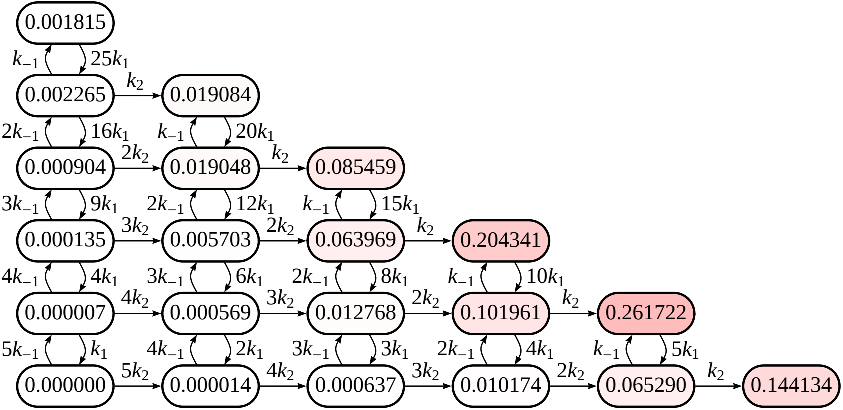


Transient probability, $t = 110$

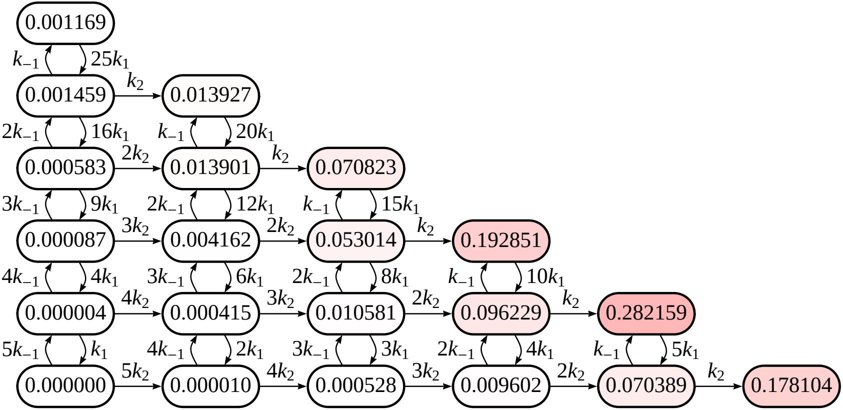
CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH



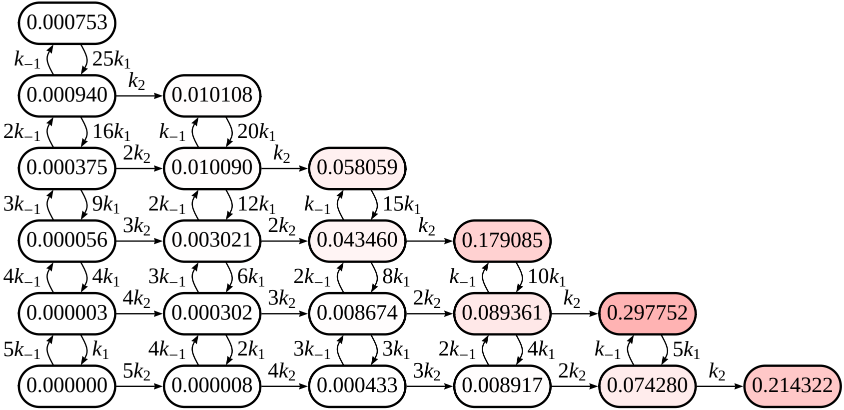
Transient probability, $t = 120$



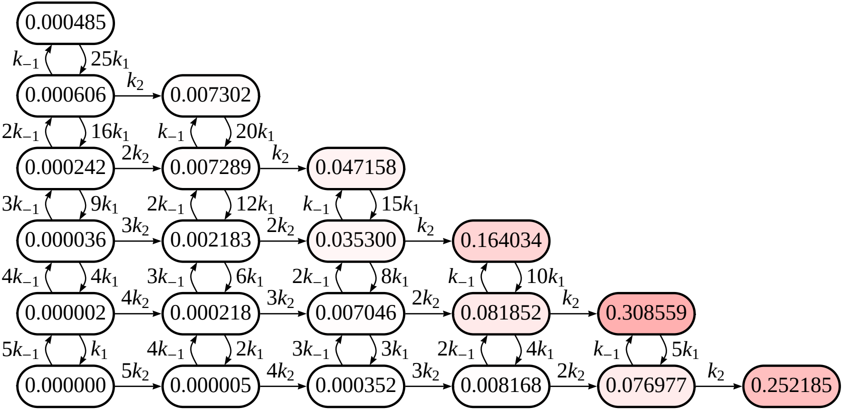
Transient probability, $t = 130$



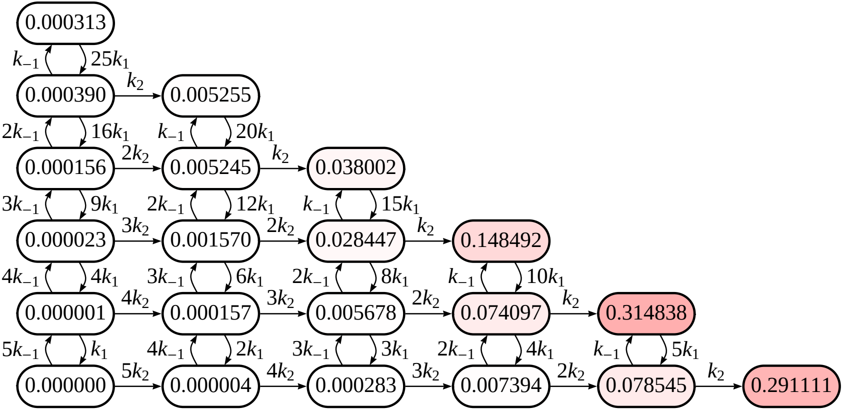
Transient probability, $t = 140$



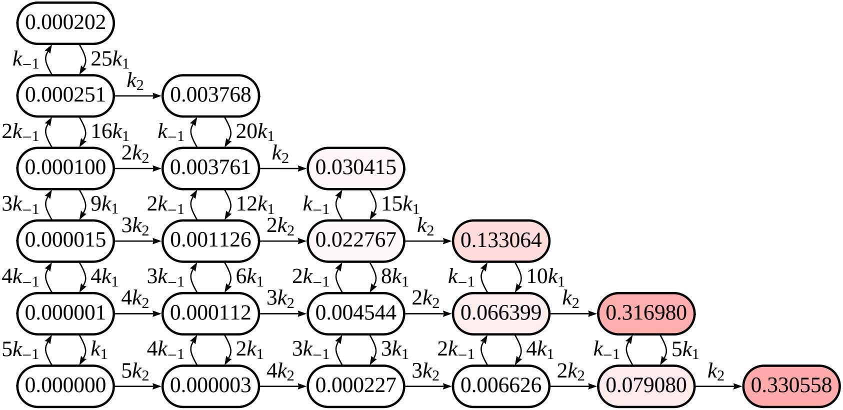
Transient probability, $t = 150$



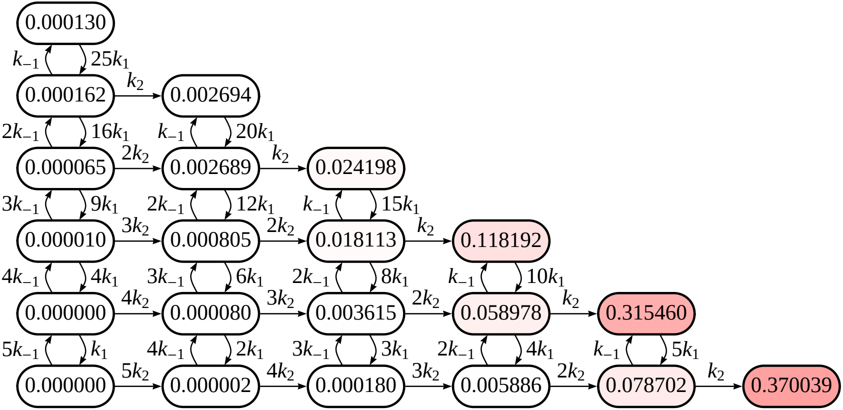
Transient probability, $t = 160$

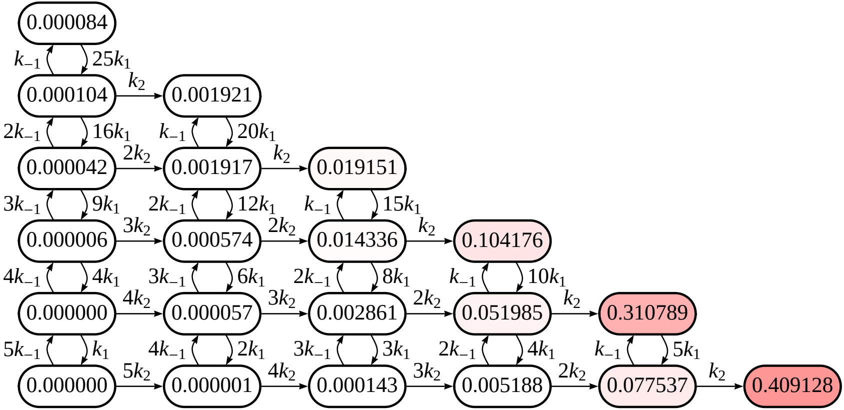


Transient probability, $t = 170$

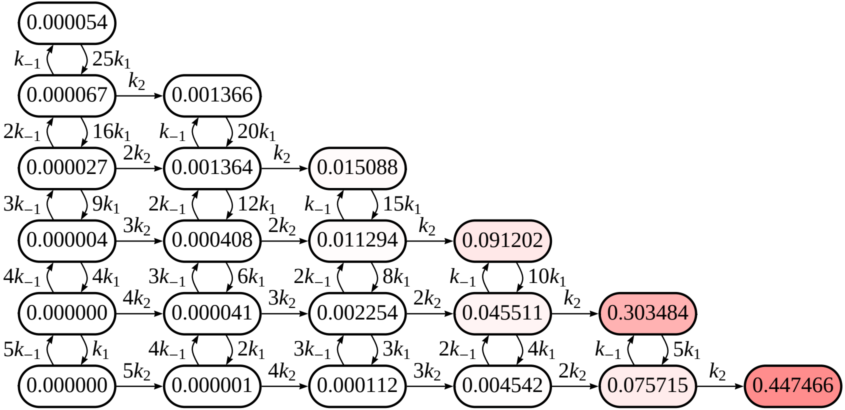


Transient probability, $t = 180$

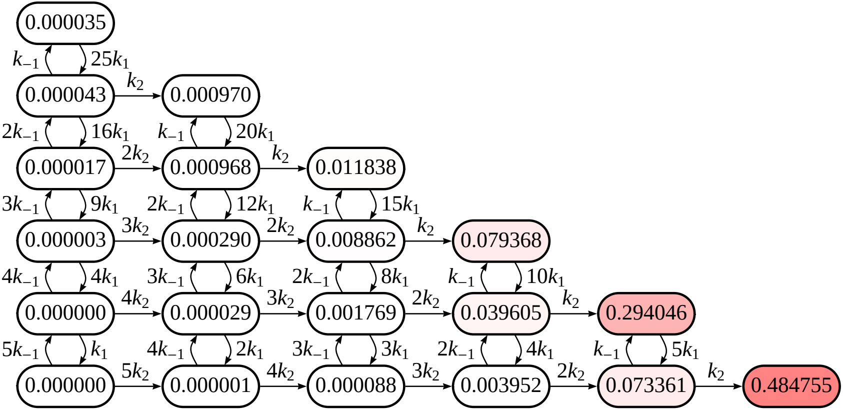



0.000084
k_{-1}
$25k_1$
0.000104
k_2
0.001921
$2k_{-1}$
$16k_1$
k_{-1}
$20k_1$
0.000042
$2k_2$
0.001917
k_2
0.019151
$3k_{-1}$
$9k_1$
$2k_{-1}$
$12k_1$
k_{-1}
$15k_1$
0.000006
$3k_2$
0.000574
$2k_2$
0.014336
k_2
0.104176
$4k_{-1}$
$4k_1$
$3k_{-1}$
$6k_1$
$2k_{-1}$
$8k_1$
k_{-1}
$10k_1$
0.000000
$4k_2$
0.000057
$3k_2$
0.002861
$2k_2$
0.051985
k_2
0.310789
$5k_{-1}$
k_1
$4k_{-1}$
$2k_1$
$3k_{-1}$
$3k_1$
$2k_{-1}$
$4k_1$
k_{-1}
$5k_1$
0.000000
$5k_2$
0.000001
$4k_2$
0.000143
$3k_2$
0.005188
$2k_2$
0.077537
k_2
0.409128

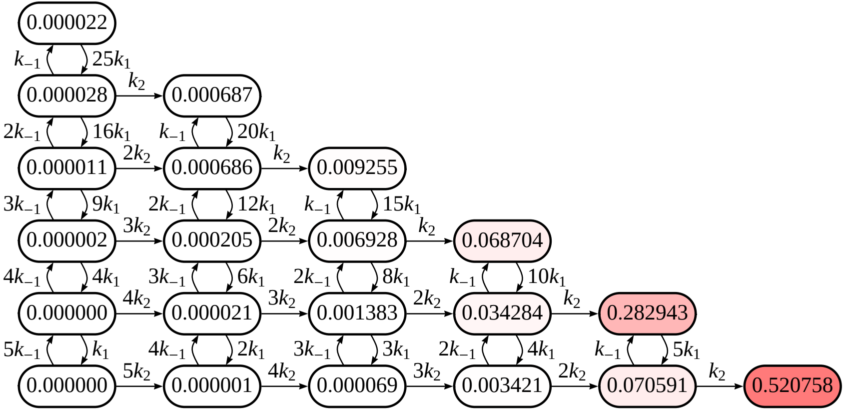
Transient probability, $t = 200$



Transient probability, $t = 210$

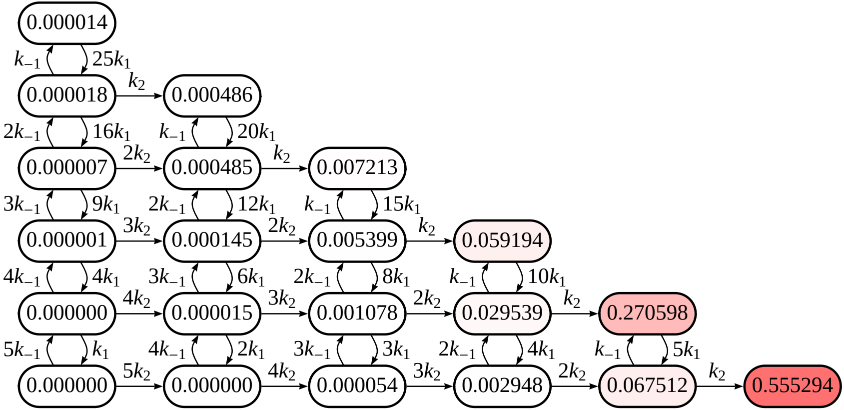


Transient probability, $t = 220$

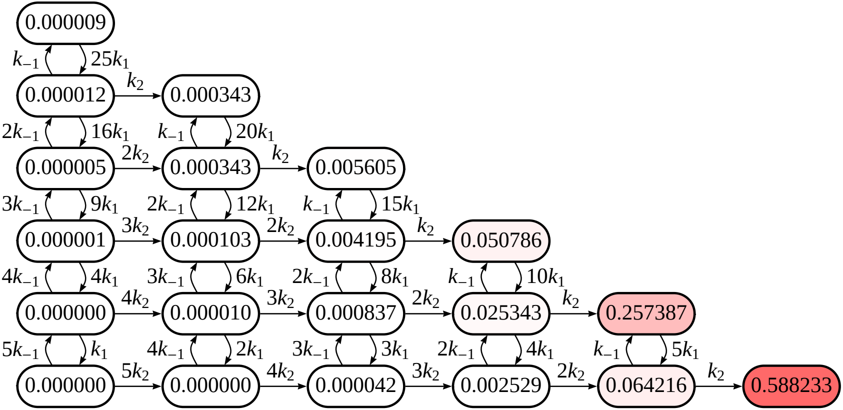


Transient probability, $t = 230$

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

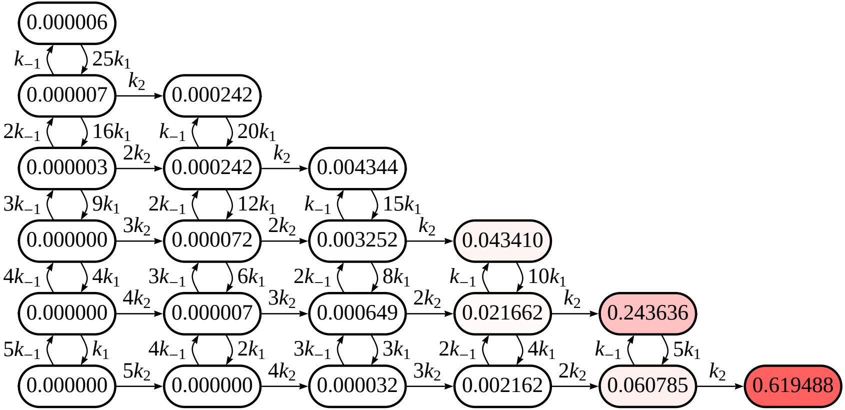


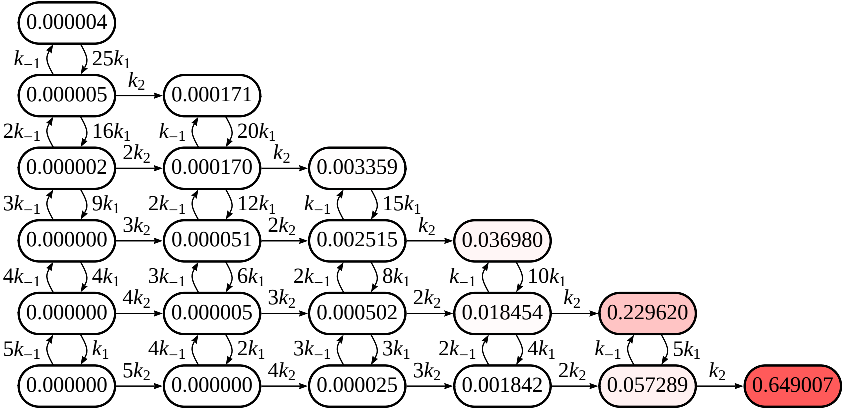
Transient probability, $t = 240$



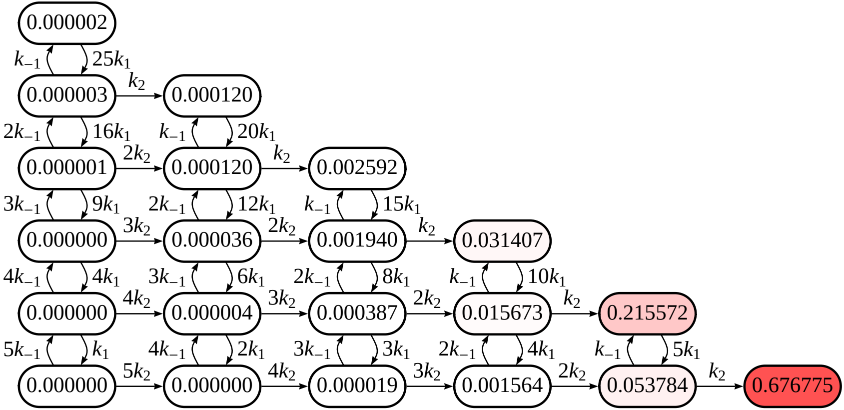
Transient probability, $t = 250$

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

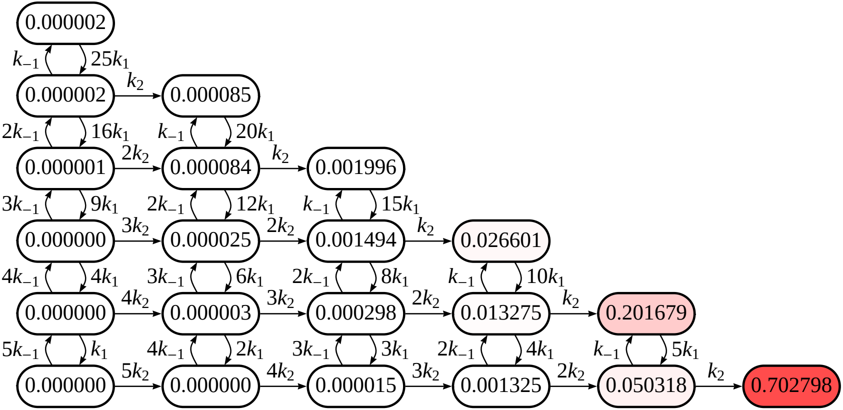


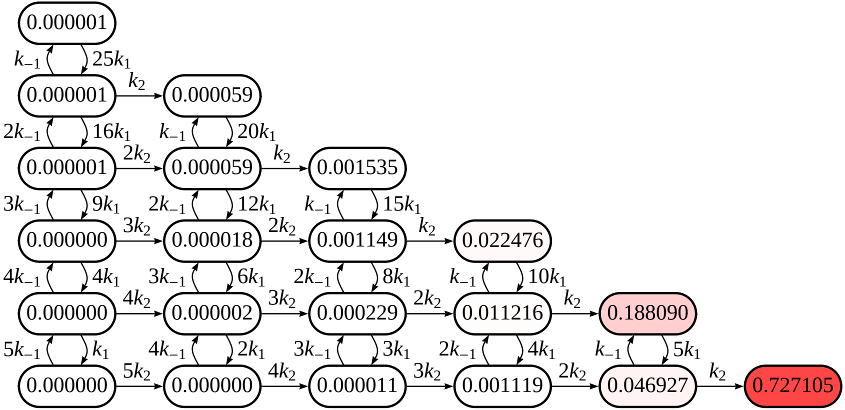

0.000004
0.000005
0.000171
0.000002
0.000170
0.003359
0.000000
0.000051
0.002515
0.036980
0.000000
0.000005
0.000502
0.018454
0.229620
0.000000
0.000000
0.000025
0.001842
0.057289
0.649007

Transient probability, $t = 270$



Transient probability, $t = 280$

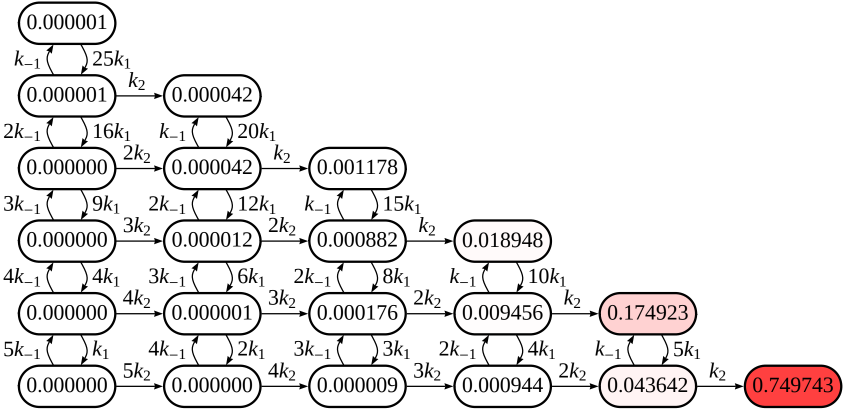




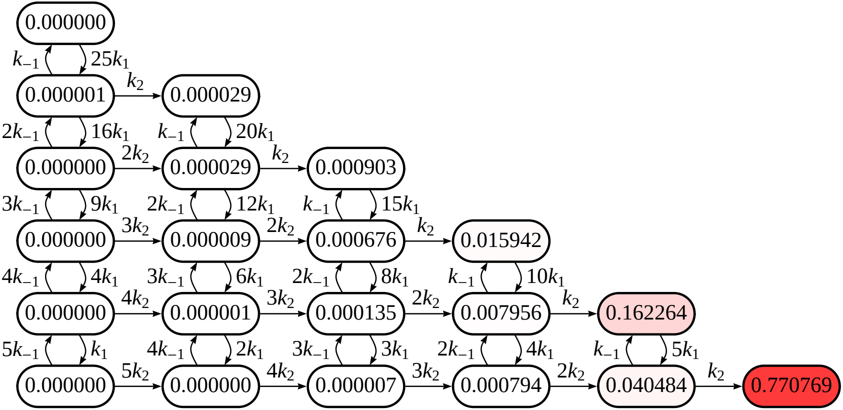
0.000001
0.000001
0.000059
0.000001
0.000059
0.001535
0.000000
0.000018
0.001149
0.022476
0.000000
0.000002
0.000229
0.011216
0.188090
0.000000
0.000000
0.000011
0.001119
0.046927
0.727105
k_{-1}
$25k_1$
k_2
$2k_{-1}$
$16k_1$
k_{-1}
$20k_1$
$2k_2$
k_2
$3k_{-1}$
$9k_1$
$2k_{-1}$
$12k_1$
k_{-1}
$15k_1$
$3k_2$
$2k_2$
k_2
$4k_{-1}$
$4k_1$
$3k_{-1}$
$6k_1$
$2k_{-1}$
$8k_1$
k_{-1}
$10k_1$
$4k_2$
$3k_2$
$2k_2$
k_2
$5k_{-1}$
k_1
$4k_{-1}$
$2k_1$
$3k_{-1}$
$3k_1$
$2k_{-1}$
$4k_1$
k_{-1}
$5k_1$
$5k_2$
$4k_2$
$3k_2$
$2k_2$
k_2

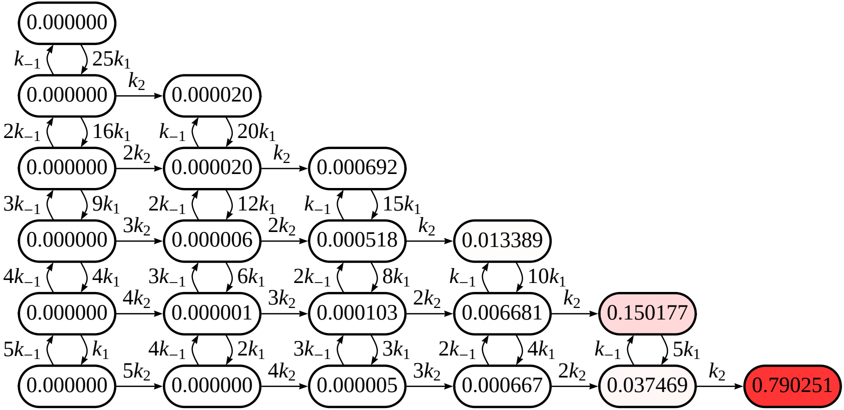
Transient probability, $t = 300$

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

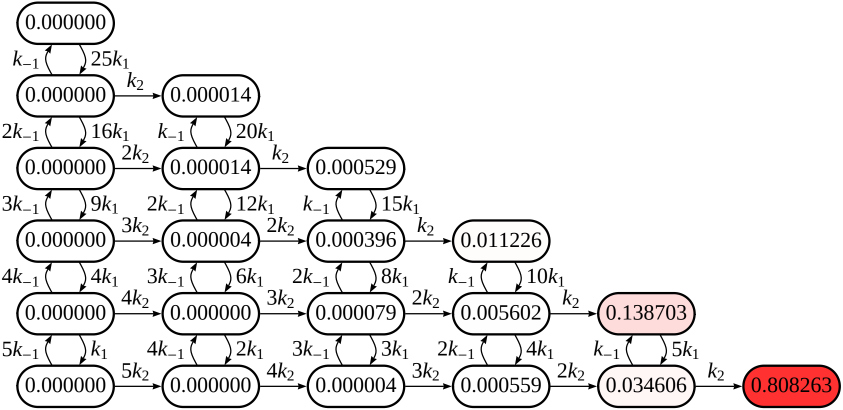


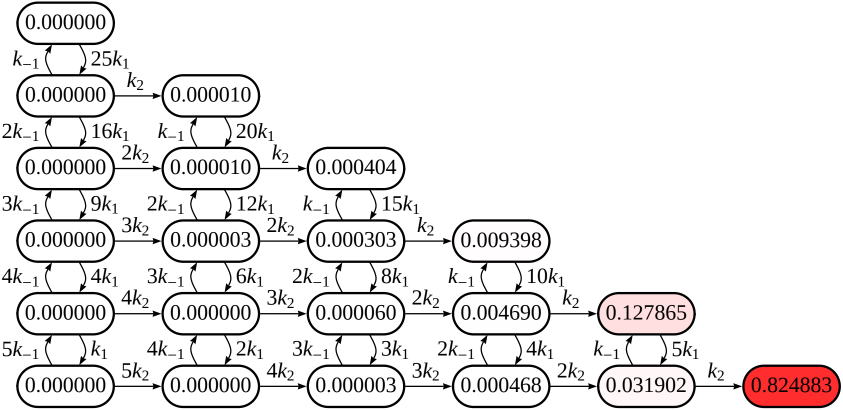
Transient probability, $t = 310$



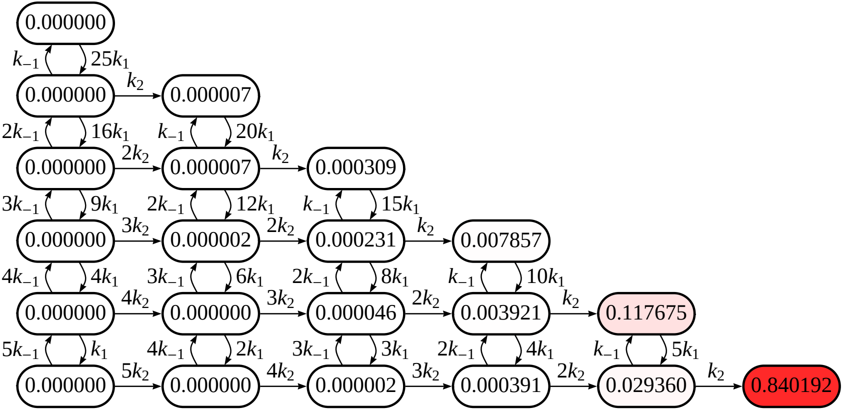

0.000000
0.000000
0.000020
0.000000
0.000020
0.000692
0.000000
0.000006
0.000518
0.013389
0.000000
0.000001
0.000103
0.006681
0.150177
0.000000
0.000000
0.000005
0.000667
0.037469
0.790251
k_{-1}
$25k_1$
k_2
$2k_{-1}$
$16k_1$
k_{-1}
$20k_1$
$2k_2$
$3k_{-1}$
$9k_1$
$2k_{-1}$
$12k_1$
k_{-1}
$15k_1$
$3k_2$
$4k_{-1}$
$4k_1$
$3k_{-1}$
$6k_1$
$2k_{-1}$
$8k_1$
$10k_1$
$4k_2$
$5k_{-1}$
k_1
$4k_{-1}$
$2k_1$
$3k_{-1}$
$3k_1$
$4k_1$
$5k_1$
$5k_2$

Transient probability, $t = 330$

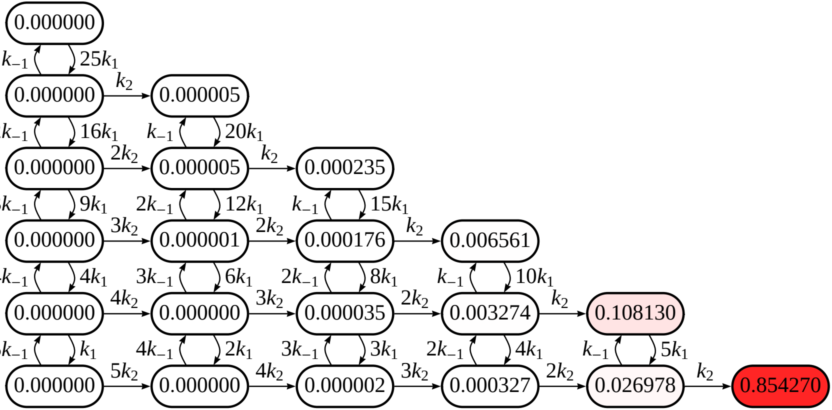



0.000000
k_{-1}
$25k_1$
0.000000
k_2
0.000010
$2k_{-1}$
$16k_1$
k_{-1}
$20k_1$
0.000000
$2k_2$
0.000010
k_2
0.000404
$3k_{-1}$
$9k_1$
$2k_{-1}$
$12k_1$
k_{-1}
$15k_1$
0.000000
$3k_2$
0.000003
$2k_2$
0.000303
k_2
0.009398
$4k_{-1}$
$4k_1$
$3k_{-1}$
$6k_1$
$2k_{-1}$
$8k_1$
k_{-1}
$10k_1$
0.000000
$4k_2$
0.000000
$3k_2$
0.000060
$2k_2$
0.004690
k_2
0.127865
$5k_{-1}$
k_1
$4k_{-1}$
$2k_1$
$3k_{-1}$
$3k_1$
$2k_{-1}$
$4k_1$
k_{-1}
$5k_1$
0.000000
$5k_2$
0.000000
$4k_2$
0.000003
$3k_2$
0.000468
$2k_2$
0.031902
k_2
0.824883

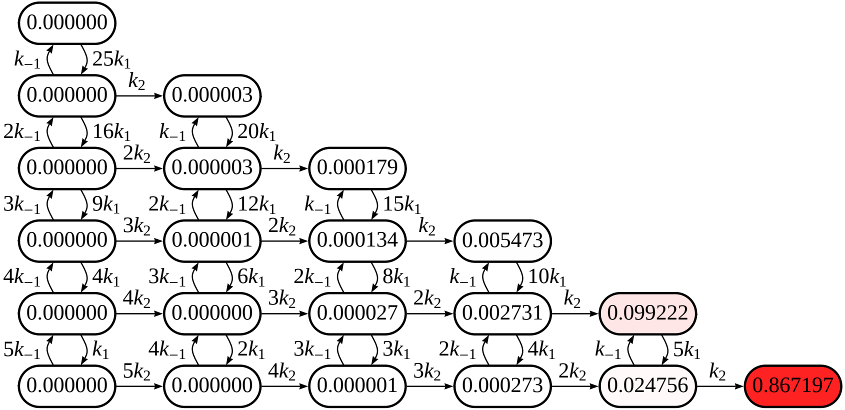
Transient probability, $t = 350$



Transient probability, $t = 360$

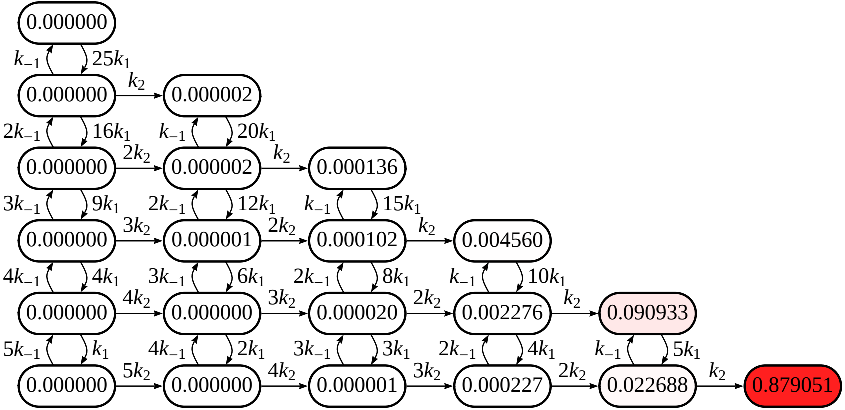


Transient probability, $t = 370$

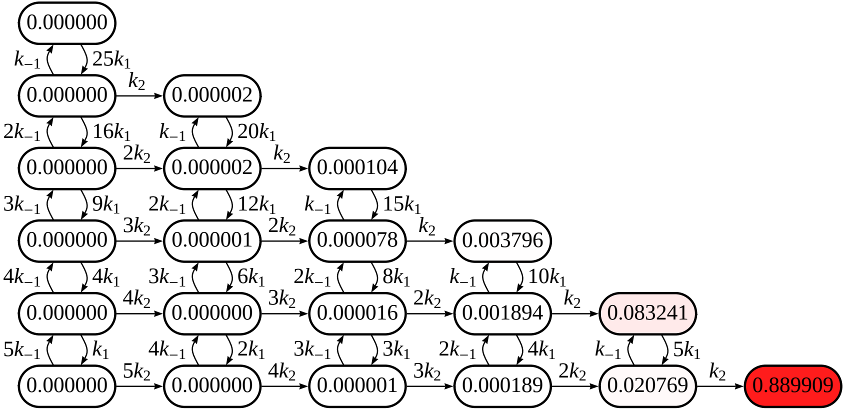


Transient probability, $t = 380$

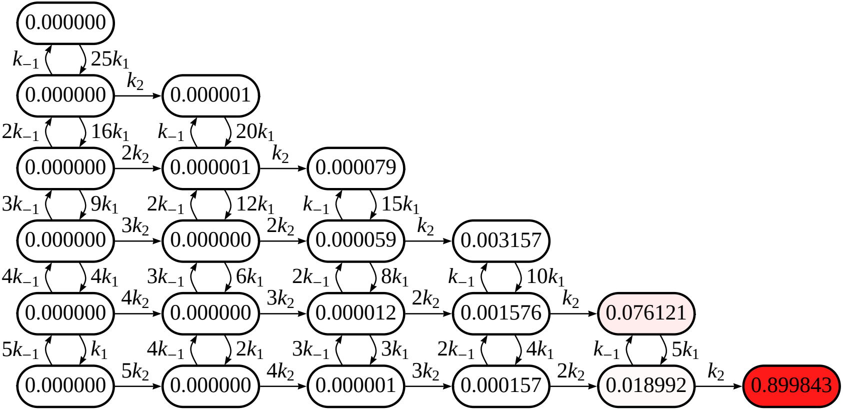
CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH



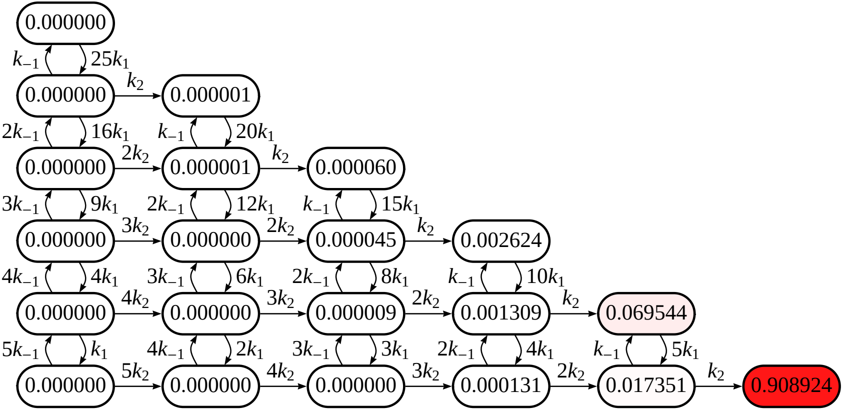
Transient probability, $t = 390$



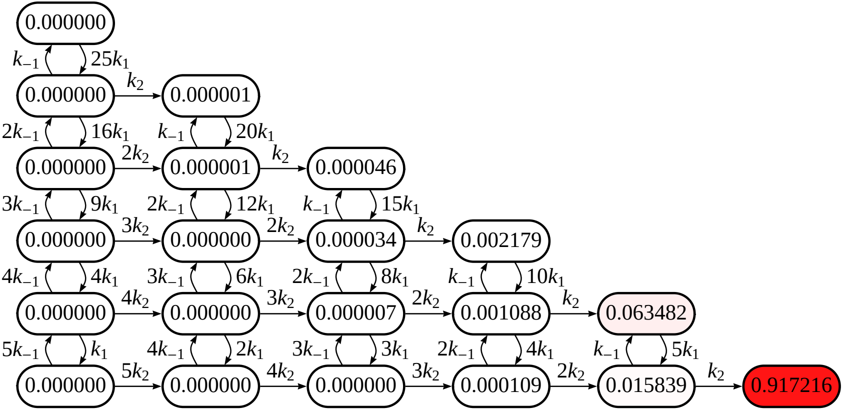
Transient probability, $t = 400$



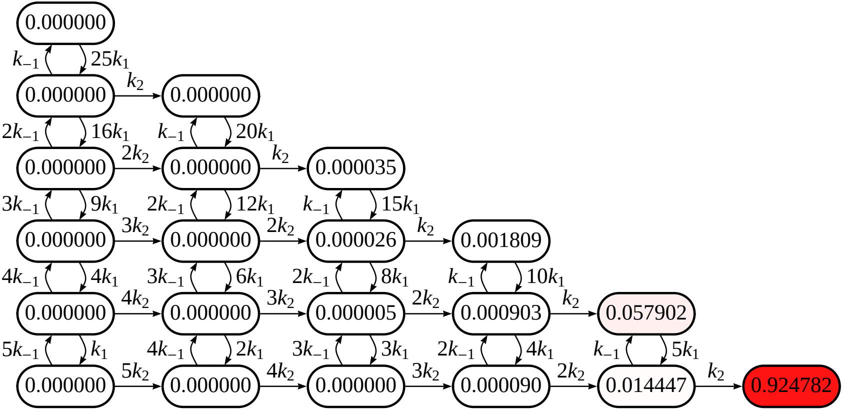
Transient probability, $t = 410$



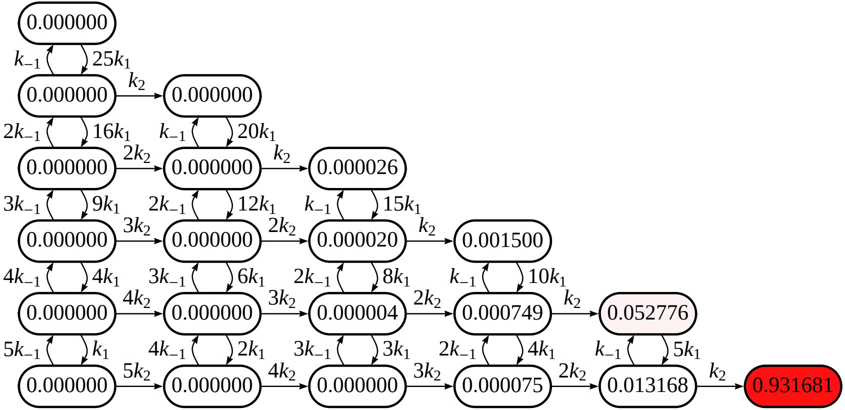
Transient probability, $t = 420$



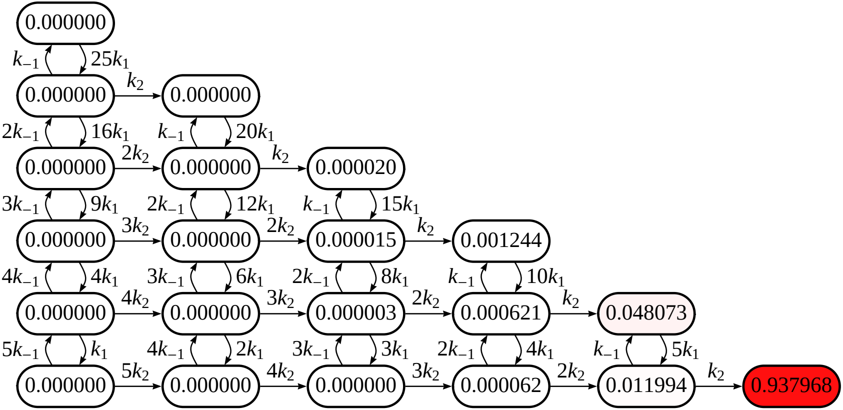
Transient probability, $t = 430$



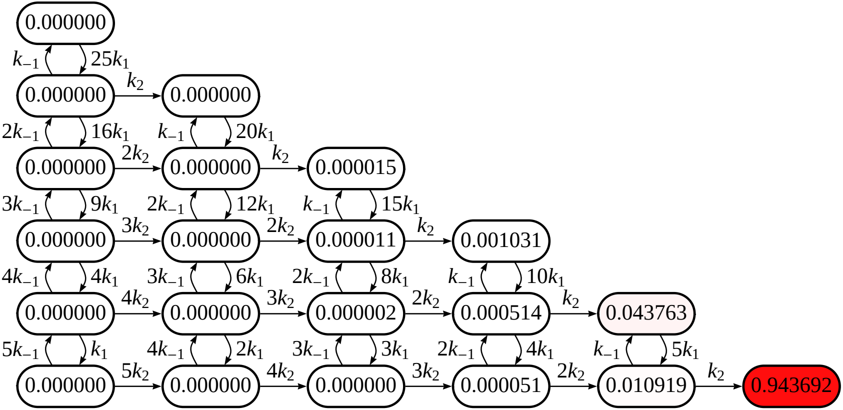
Transient probability, $t = 440$



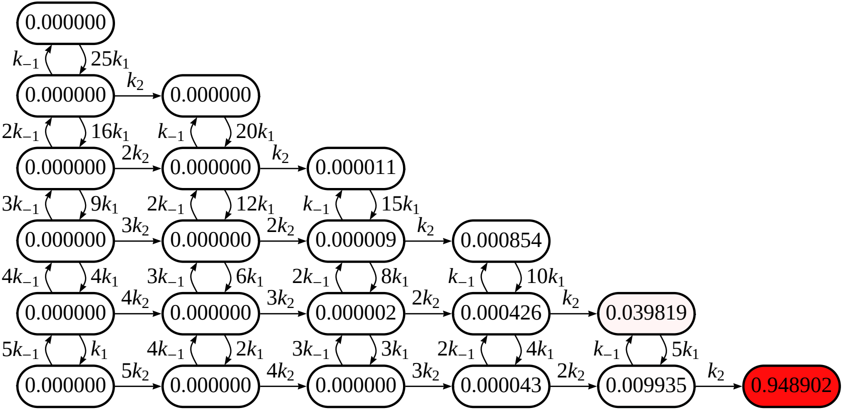
Transient probability, $t = 450$

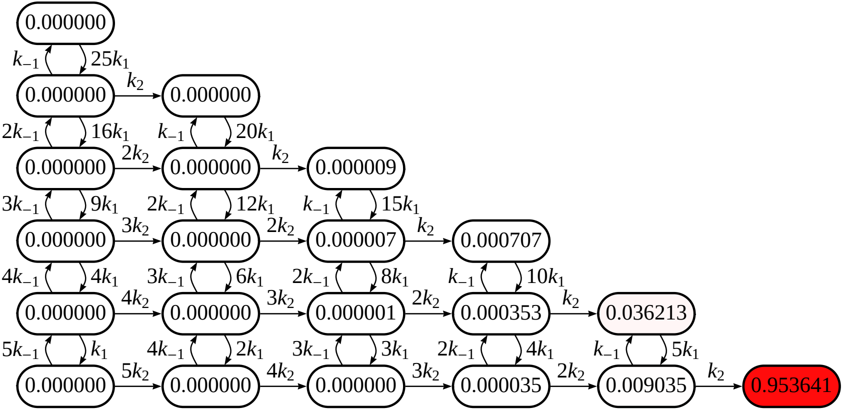


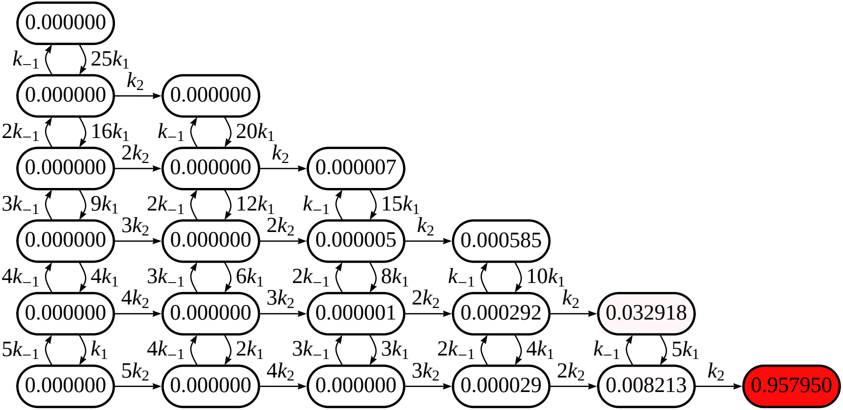
Transient probability, $t = 460$



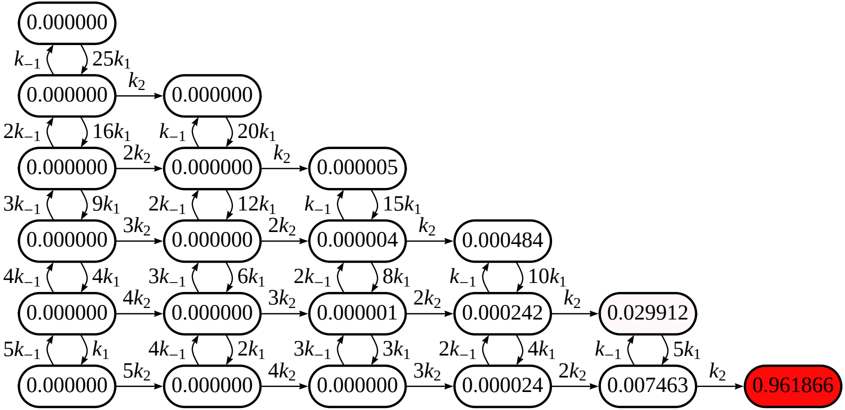
Transient probability, $t = 470$

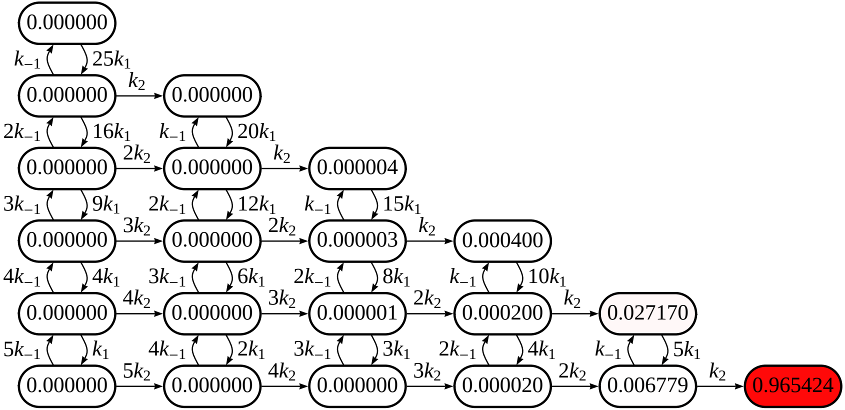



0.000000
k_{-1} $25k_1$
0.000000 k_2 0.000000
$2k_{-1}$ $16k_1$ k_{-1} $20k_1$
0.000000 $2k_2$ 0.000000 k_2 0.000009
$3k_{-1}$ $9k_1$ $2k_{-1}$ $12k_1$ k_{-1} $15k_1$
0.000000 $3k_2$ 0.000000 $2k_2$ 0.000007 k_2 0.000707
$4k_{-1}$ $4k_1$ $3k_{-1}$ $6k_1$ $2k_{-1}$ $8k_1$ k_{-1} $10k_1$
0.000000 $4k_2$ 0.000000 $3k_2$ 0.000001 $2k_2$ 0.000353 k_2 0.036213
$5k_{-1}$ k_1 $4k_{-1}$ $2k_1$ $3k_{-1}$ $3k_1$ $2k_{-1}$ $4k_1$ k_{-1} $5k_1$
0.000000 $5k_2$ 0.000000 $4k_2$ 0.000000 $3k_2$ 0.000035 $2k_2$ 0.009035 k_2 0.953641

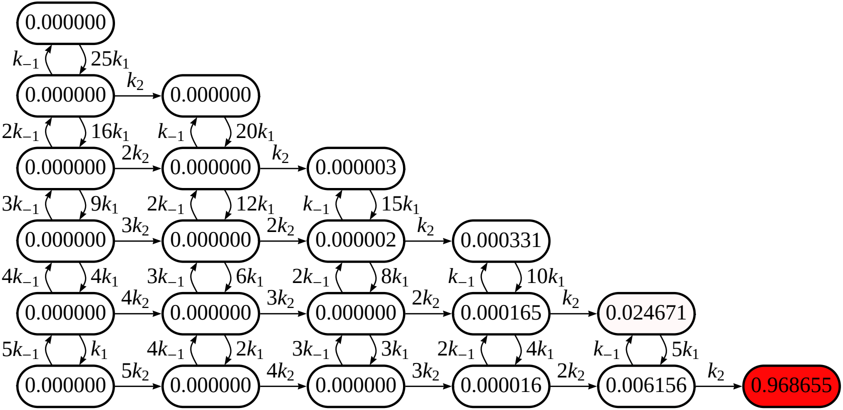

0.000000
k_{-1}
$25k_1$
0.000000
k_2
0.000000
$2k_{-1}$
$16k_1$
k_{-1}
$20k_1$
0.000000
$2k_2$
0.000000
k_2
0.000007
$3k_{-1}$
$9k_1$
$2k_{-1}$
$12k_1$
k_{-1}
$15k_1$
0.000000
$3k_2$
0.000000
$2k_2$
0.000005
k_2
0.000585
$4k_{-1}$
$4k_1$
$3k_{-1}$
$6k_1$
$2k_{-1}$
$8k_1$
k_{-1}
$10k_1$
0.000000
$4k_2$
0.000000
$3k_2$
0.000001
$2k_2$
0.000292
k_2
0.032918
$5k_{-1}$
k_1
$4k_{-1}$
$2k_1$
$3k_{-1}$
$3k_1$
$2k_{-1}$
$4k_1$
k_{-1}
$5k_1$
0.000000
$5k_2$
0.000000
$4k_2$
0.000000
$3k_2$
0.000029
$2k_2$
0.008213
k_2
0.957950

Transient probability, $t = 500$

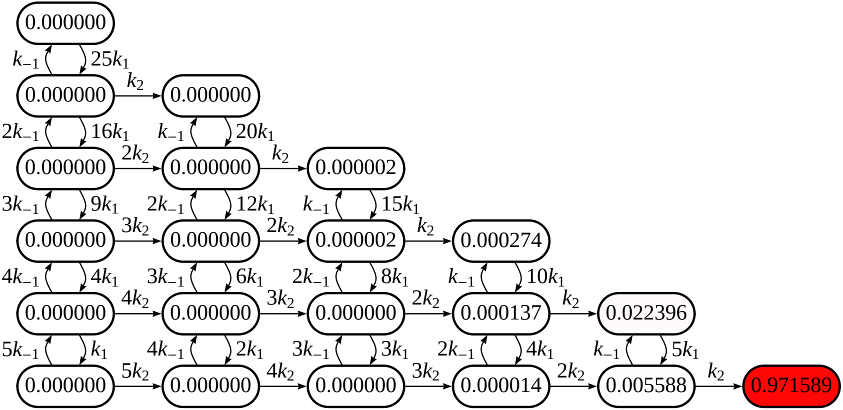



0.000000
0.000000
0.000000
0.000000
0.000000
0.000004
0.000000
0.000000
0.000003
0.000400
0.000000
0.000000
0.000001
0.000200
0.027170
0.000000
0.000000
0.000000
0.000020
0.006779
0.965424
k_{-1}
$25k_1$
k_2
$2k_{-1}$
$16k_1$
k_{-1}
$20k_1$
$2k_2$
$3k_{-1}$
$9k_1$
$2k_{-1}$
$12k_1$
k_{-1}
$15k_1$
$3k_2$
$4k_{-1}$
$4k_1$
$3k_{-1}$
$6k_1$
$2k_{-1}$
$8k_1$
$10k_1$
$4k_2$
$5k_{-1}$
k_1
$4k_{-1}$
$2k_1$
$3k_{-1}$
$3k_1$
$4k_1$
$5k_1$
$5k_2$

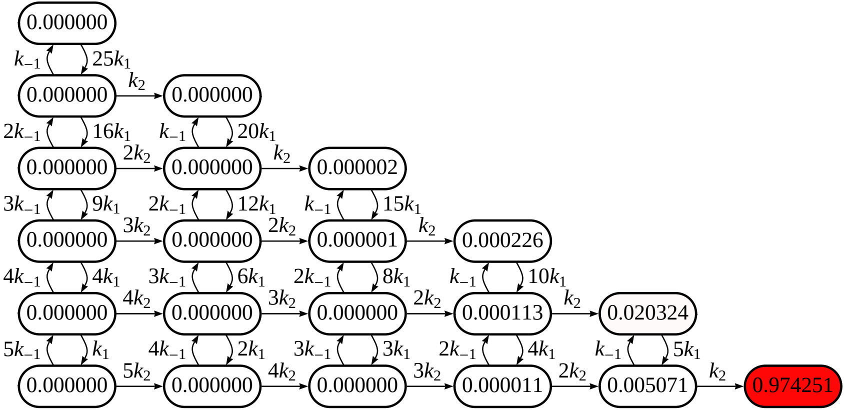
Transient probability, $t = 520$



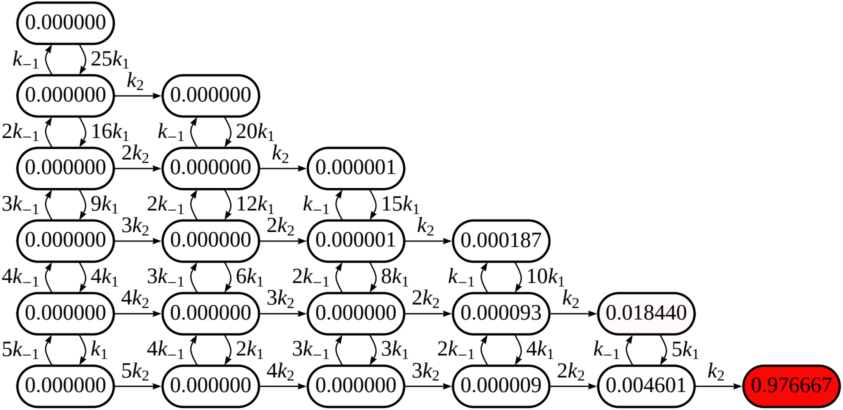
Transient probability, $t = 530$



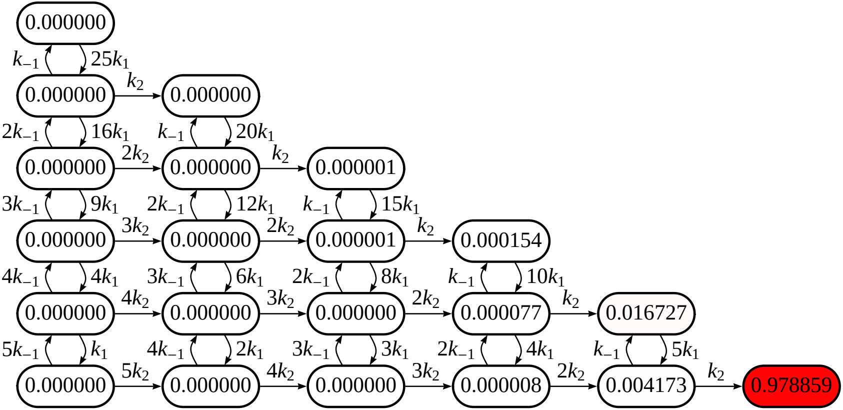
Transient probability, $t = 540$

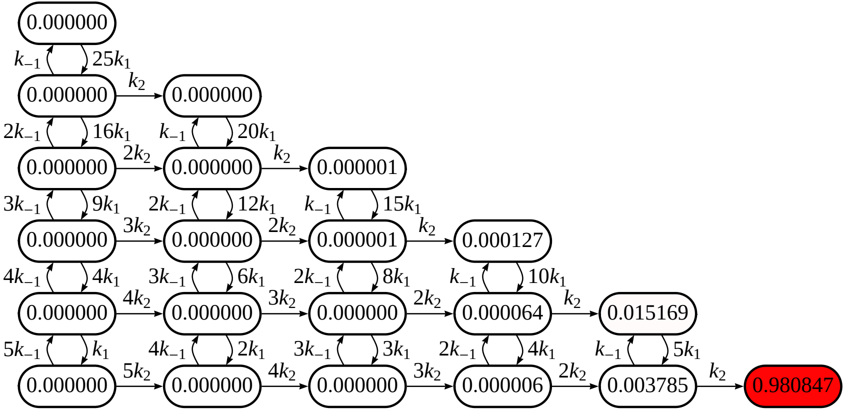


Transient probability, $t = 550$

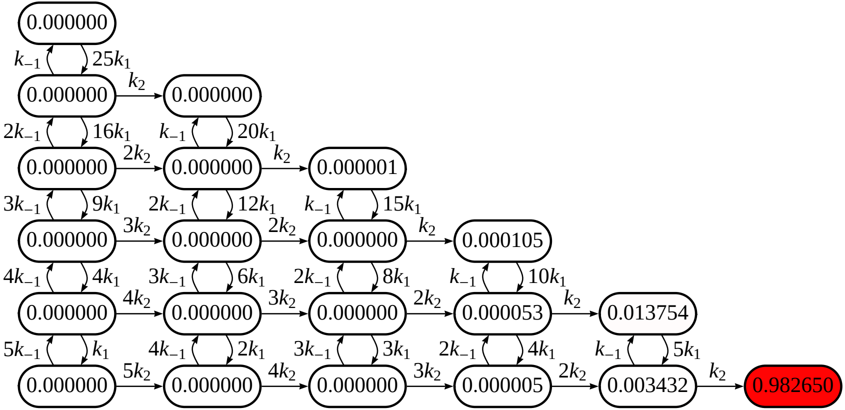


Transient probability, $t = 560$

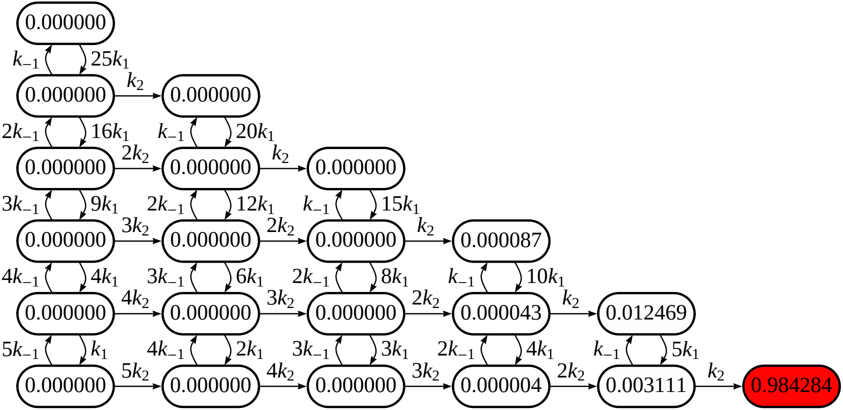



0.000000
k_{-1}
$25k_1$
0.000000
k_2
0.000000
$2k_{-1}$
$16k_1$
k_{-1}
$20k_1$
0.000000
$2k_2$
0.000000
k_2
0.000001
$3k_{-1}$
$9k_1$
$2k_{-1}$
$12k_1$
k_{-1}
$15k_1$
0.000000
$3k_2$
0.000000
$2k_2$
0.000001
k_2
0.000127
$4k_{-1}$
$4k_1$
$3k_{-1}$
$6k_1$
$2k_{-1}$
$8k_1$
k_{-1}
$10k_1$
0.000000
$4k_2$
0.000000
$3k_2$
0.000000
$2k_2$
0.000064
k_2
0.015169
$5k_{-1}$
k_1
$4k_{-1}$
$2k_1$
$3k_{-1}$
$3k_1$
$2k_{-1}$
$4k_1$
k_{-1}
$5k_1$
0.000000
$5k_2$
0.000000
$4k_2$
0.000000
$3k_2$
0.000006
$2k_2$
0.003785
k_2
0.980847

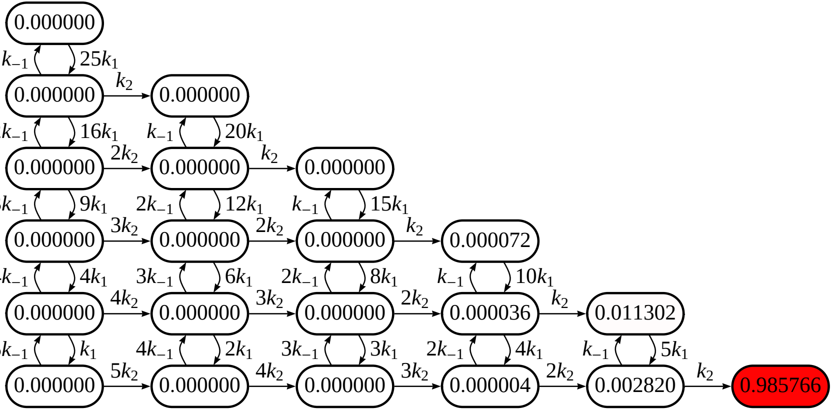
Transient probability, $t = 580$



Transient probability, $t = 590$



Transient probability, $t = 600$

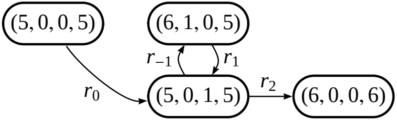


- If we consider an extension of the model with an additional reaction r_0 which synthesises the compound $E{:}S$ as shown below with the synthesis occurring at a constant rate $r_0 = k_0$ then this additional reaction channel changes the analysis of the model dramatically.

$E{:}S$	$r_0 \uparrow \quad + \quad r_1 \uparrow \quad + \quad r_{-1} \downarrow \quad + \quad r_2 \downarrow$

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

- The state which was previously a deadlock state now admits an r_0 reaction which leads it to a previously unreachable state, $(5, 0, 1, 5)$. The reactions r_{-1}, r_1 and r_2 can occur in states reachable from that.



- Each of these states, *and every other state*, now allows an r_0 reaction, taking them to previously unreachable states each of which allows r_0 and reactions r_{-1}, r_1 and r_2 subsequent to that.

Adding synthesis

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

- Each of these states, *and every other state*, now allows an r_0 reaction, taking them to previously unreachable states each of which allows r_0 and reactions r_{-1}, r_1 and r_2 subsequent to that.
- The effect of introducing this single synthesis reaction is that we now cannot find any upper bound N such that the molecular species counts are guaranteed to lie in the bounded integer range 0 to N.

Adding synthesis

- Each of these states, *and every other state*, now allows an r_0 reaction, taking them to previously unreachable states each of which allows r_0 and reactions r_{-1}, r_1 and r_2 subsequent to that.
- The effect of introducing this single synthesis reaction is that we now cannot find any upper bound N such that the molecular species counts are guaranteed to lie in the bounded integer range 0 to N.
- If we are unable to bound the reachable state-space then we cannot in general analyse our model by probabilistic model-checking.

1. The generation of the derivation graph of the underlying state-space does not take into account the numerical values assigned to the rate constants, and the propensity functions which depend on those. This means that the derivation graph may include many states which the system is almost sure not to reach within a particular time bound.

2. Most chemical systems involve several widely varying time scales, so such systems are nearly always stiff. A consequence of this is that the first passage time to many states is likely to be long and truncation of the state-space using a time-bounded reachability metric is likely to be productive.

3. Many of the logical formulae which we wish to check involve reaching within a fixed time bound model states which satisfy a given predicate.

4. Stochastic simulation methods such as Gillespie's Direct Method generate exact stochastic simulations of trajectories from the initial state to states reachable within a given time bound.

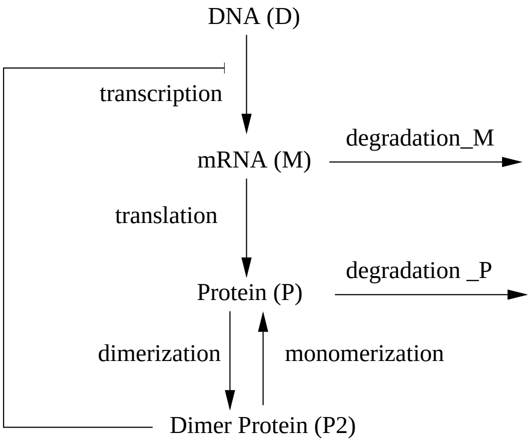
Outline

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

Examples: Two Genetic Networks

- In order to illustrate our approach we consider two models. These represent, under different assumptions, a general genetic network with a negative feedback. An example of this kind of network is the control circuit for the λ repressor protein CI of λ-phage in *E.Coli*.
- We have four biochemical entities that interact with each other through six reactions. The biochemical entities are the DNA (D), the mRNA (M), a protein in monomeric form (P) and a protein in dimeric form ($P2$).

A schema of the general network



The network is unbounded

CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH

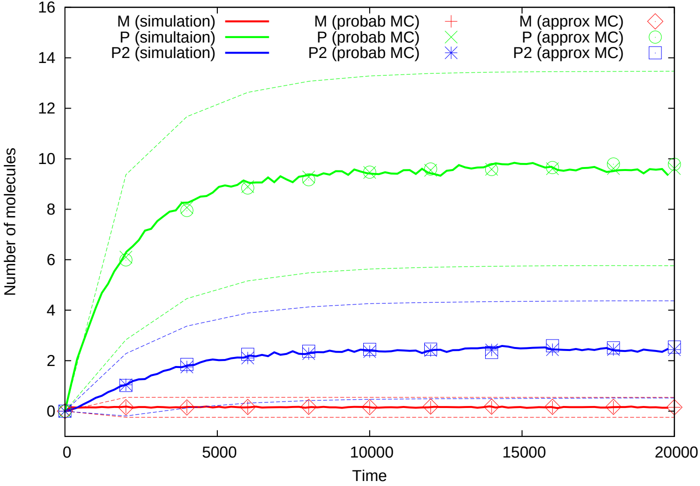
- The network is structurally unbounded, since both transcription and translation lead to the creation of new molecules.
- However, the two degradation reactions and the transcription inhibition by means of the dimeric protein have a regulatory effect on the protein synthesis and therefore, under some conditions, all the species reach a finite average value.

- We perform 1000 independent stochastic simulation runs using Gillespie's Direct Method. The number of runs is large enough to take into account the variability of the system, but still making the total simulation time reasonable. We used $T = 20000\ s$ as a simulation stop time: by that time the system has reached a stable state.

- We can estimate the upper bounds for the amounts of each species as the maximum values obtained in any run at any time instant, and we can use these values in the PRISM model.

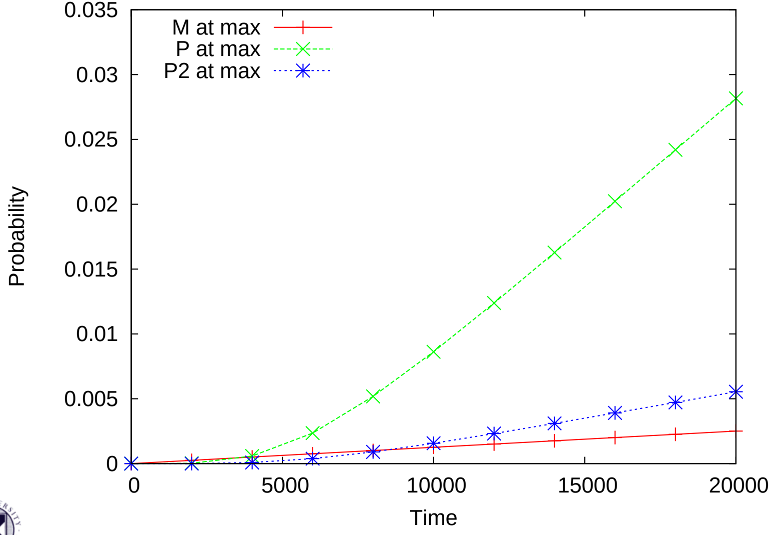
$$Max_M = 5; \quad Max_P = 33; \quad Max_{P2} = 18$$

Simulation averages and model-checking for $\mathcal{M}_1$



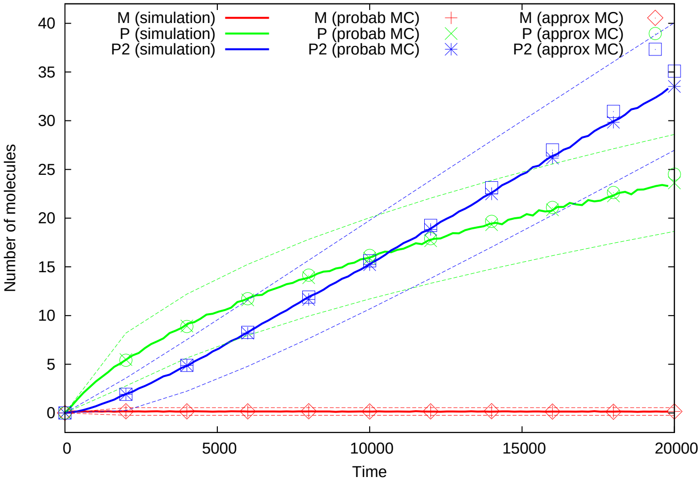
- As another form of validation of the derived bounds, we have calculated the probabilities of reaching them at different time instants:
  - $\mathcal{P}_{=?}[\mathit{true}\ \mathrm{U}^{\leq T}\ M = 5]$,
  - $\mathcal{P}_{=?}[\mathit{true}\ \mathrm{U}^{\leq T}\ P = 33]$, and
  - $\mathcal{P}_{=?}[\mathit{true}\ \mathrm{U}^{\leq T}\ P2 = 18]$.
- The results provide a means of estimating the error which might have been introduced by bounding the system.

Estimating the error introduced by truncation

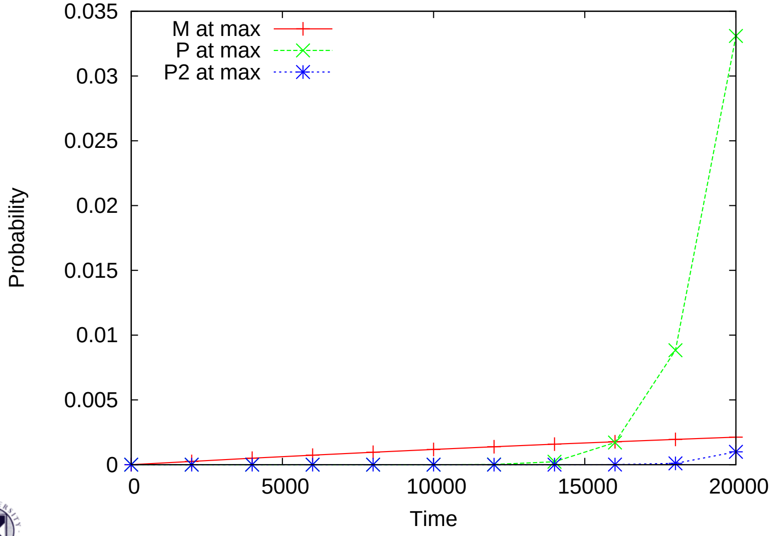


The network without protein degradation ($\mathcal{M}_2$)

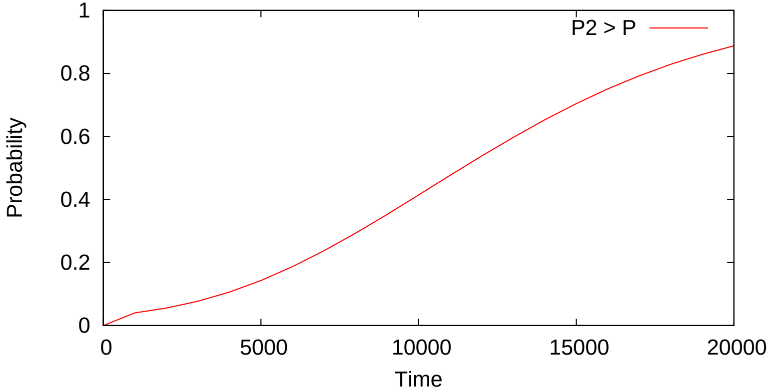
CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH



Estimating the error introduced by truncation



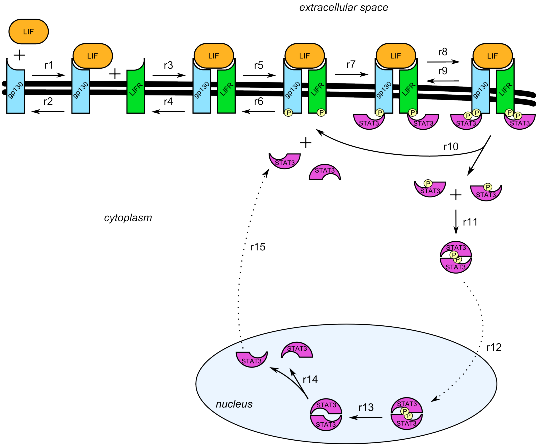
Determining the probability that $P2 > P$



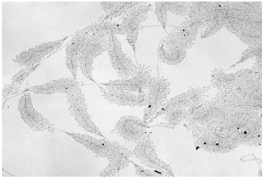
Outline

gp130/JAK/STAT pathway

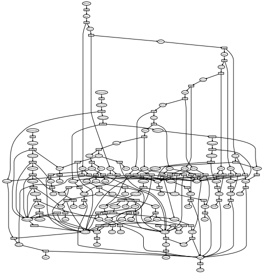
CENTRE FOR SYSTEMS BIOLOGY AT EDINBURGH



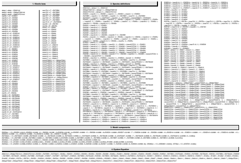
Co-transcriptional cleavage



rRNA transcription



Reaction graph



Bio-PEPA model

Federica Ciocchetta, Jane Hillston, Martin Kos and David Tollervey

Modelling co-transcriptional cleavage in the synthesis of yeast pre-rRNA.

Theoretical Computer Science

Volume 408, Issue 1, 17 November 2008, Pages 41-54.

Acknowledgements

- Adam Duguid is supported by the EPSRC Doctoral Training Grant EP/P501407/1.
- Allan Clark, Vashti Galpin and Stephen Gilmore are supported by the EPSRC grant EP/E031439/1 "Stochastic Process Algebra for Biochemical Signalling Pathway Analysis".
- Jane Hillston is supported by the Engineering and Physical Sciences Research Council (EPSRC) Advanced Research Fellowship and research grant EP/C543696/1 "Process Algebra Approaches to Collective Dynamics".
- Maria Luisa Guerriero and Laurence Loewe are supported by the Centre for Systems Biology at Edinburgh. The Centre for Systems Biology at Edinburgh is a Centre for Integrative Systems Biology (CISB) funded by BBSRC and EPSRC, reference BB/D019621/1.