

NANYANG
TECHNOLOGICAL
UNIVERSITY

NetLand: A comprehensive tool for simulation and visualization of transcriptional network kinetics

Jing Guo^{1,2}, Feng Lin¹, [Jie Zheng](#)^{1,3,*}

¹ Biomedical Informatics Graduate Lab, School of Computer Engineering, Nanyang Technological University, Singapore, Singapore

² Bioinformatics Institute of Singapore (BII), A*STAR, Singapore

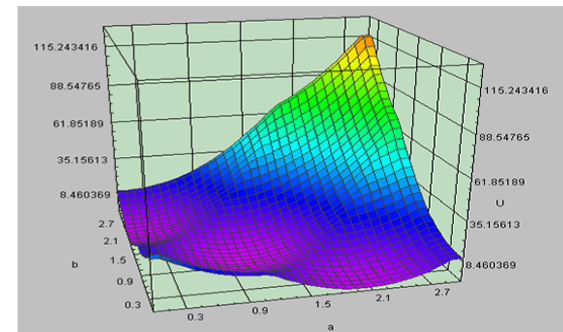
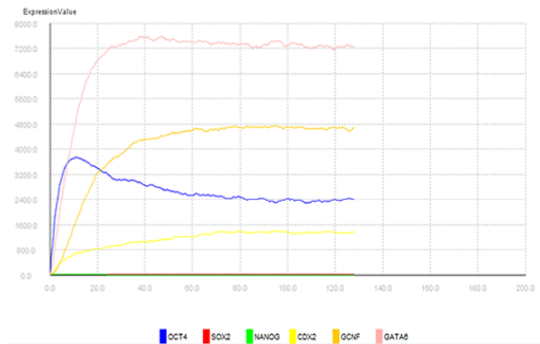
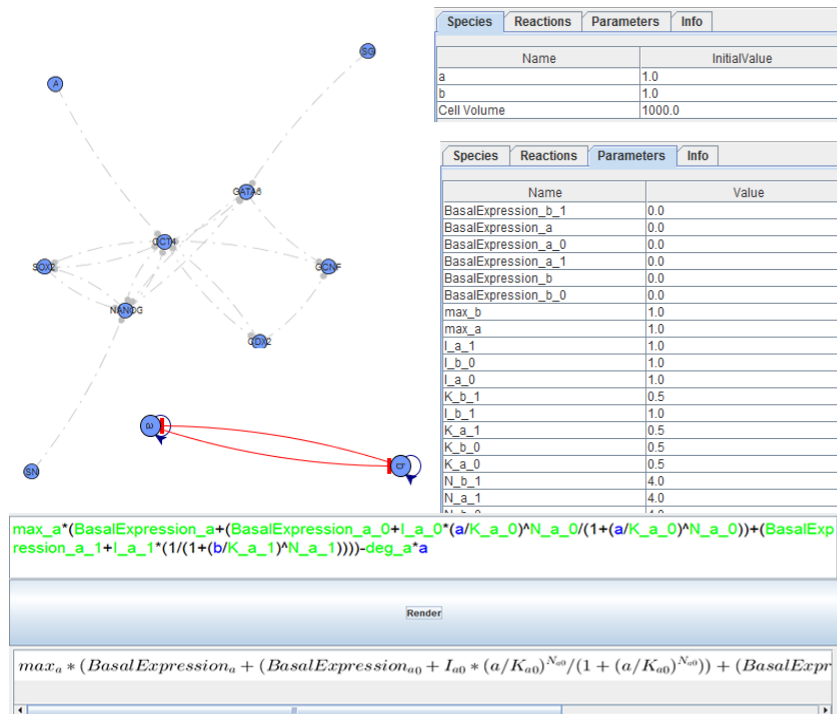
³ Genome Institute of Singapore, A*STAR, Biopolis, Singapore

* Email: ZhengJie@ntu.edu.sg

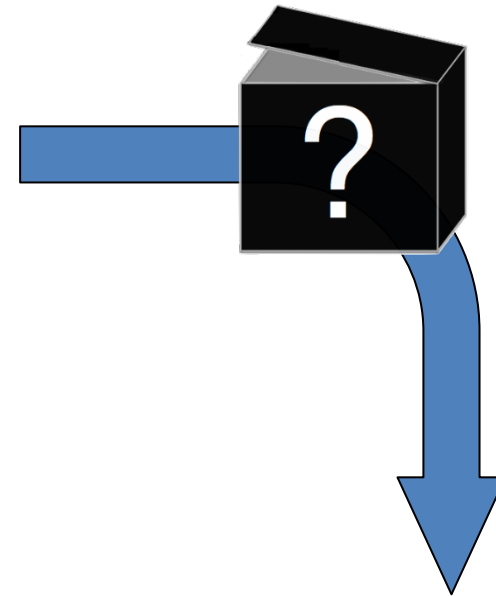
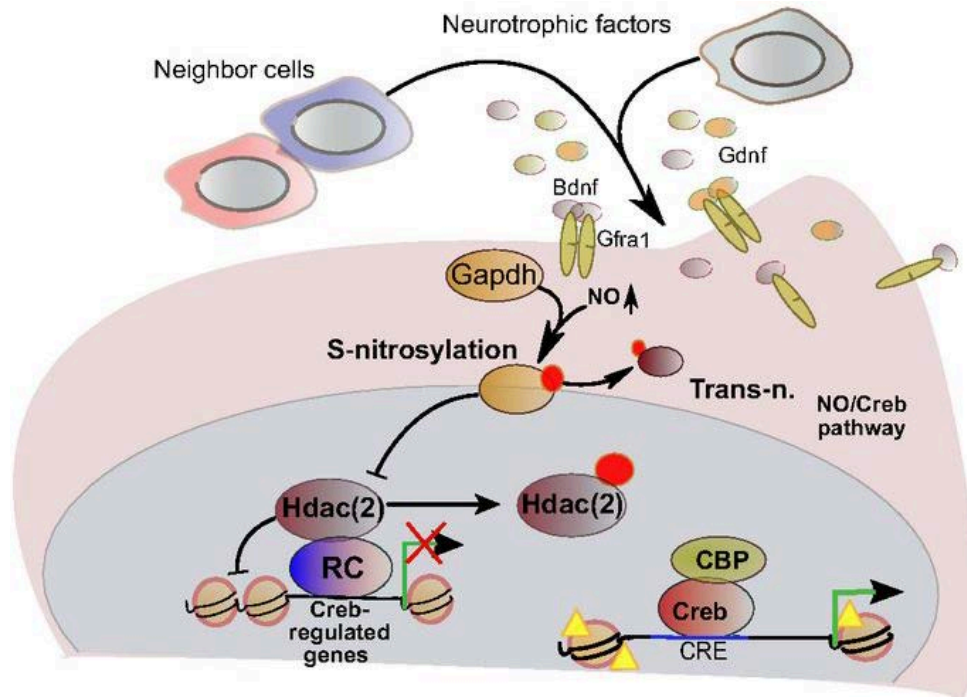
International Conference on Systems Biology (ICSB 2015)
BioNetVisA Workshop 2015
25 Nov. 2015, Singapore

NetLand – A software for global dynamics analysis on genetic system

NetLand is a standalone software tool for studying GRN kinetics. It provides comprehensive methods on simulation and visualization of network dynamics.

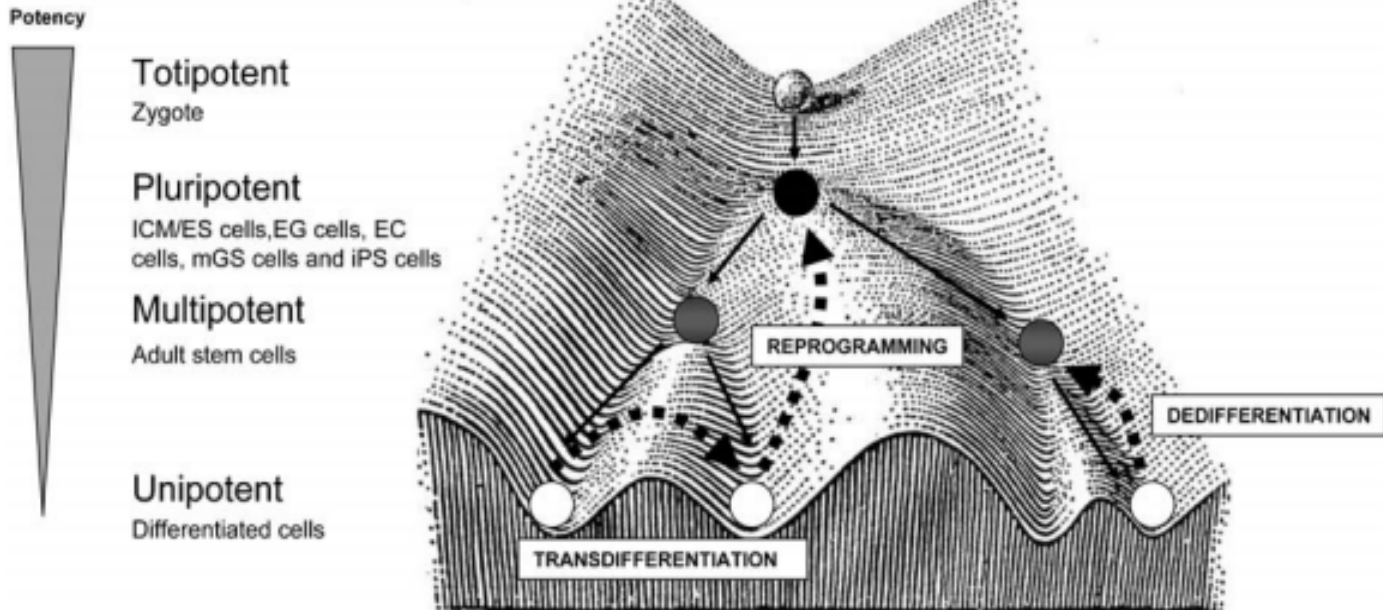


Background



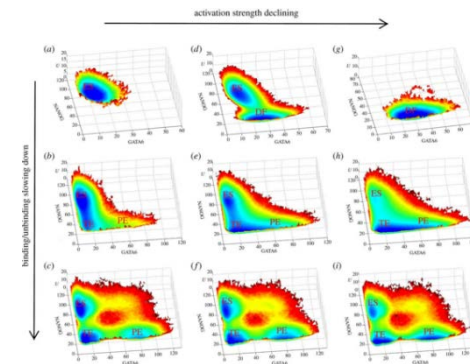
Mechanisms

- Transcriptional network
- Signaling pathway
- Other mechanisms



Differentiation, transdifferentiation and reprogramming on the **Waddington's epigenetic landscape** show trends of cellular dynamics.

The ideas borrowed from **physics** makes the qualitative metaphor a **quantitative model** which is capable of describing the global dynamics of a system.



Mathematical Formulation

Differential Equation Model

$$\frac{d[\text{Gene1}]}{dt} = K^+ \varphi(\text{Gene1}, \text{Gene2}, \text{Gene3} \dots \text{GeneN}) - K^- [\text{Gene1}] + D \cdot \theta(t)$$

Where:

- $[\text{Gene1}]$ denotes the concentration of Gene 1
- K^+ is the maximum transcriptional rate
- K^- is the degradation rate
- $\varphi()$ is kinetics of Gene 1 being regulated by other nodes (e.g. Hill function)
- $\theta(t)$ is the noise term
- D is the strength of noise

Theoretical model mainly follows:

Li C, Wang J (2013). Quantifying Cell Fate Decisions for Differentiation and Reprogramming of a Human Stem Cell Network: Landscape and Biological Paths. *PLoS Comput Biol* 9(8): e1003165.

Implement Waddington's epigenetic landscape

The probabilistic landscape is an implementation of Waddington's epigenetic landscape. It is constructed with the assumption that the noise in gene expression follows Gaussian distribution and the individual probability of each gene is independent. The quasi-potential landscape is based on the notion that the probability of gene expression states determines the stability.

Langevin dynamics that describes the movement of Brownian motion has been applied to studying the stochasticity in gene transcriptional regulation.

In genetic and biochemical circuits, the change rates of species can be formulated as Langevin equations

$$dX(t) = -\nabla U dt + \sqrt{D} dW_t$$

The number of molecules of each gene in the network at time t is denoted by vector $X(t)$. D is the noise term associated with each reaction. U is the particle interaction potential. W_t denotes a Wiener process (Standard Brownian motion). The gradient force F is equal to $-\nabla U$. The drift part is composed of the noise term.

The Langevin equation can be reformulated as a **Fokker–Planck** (FP) equation that governs the time evolution of the probability density function of each $X(t)$ under the influence of driving forces and random fluxes. FP equation is a partial differential equation (PDE).

$$\frac{\partial P(X, t)}{\partial t} = - \sum_{i=1}^N \frac{\partial [F(x_i)P(x_i, t)]}{\partial x_i} + \sum_{i=1}^N \sum_{j=1}^N \frac{\partial^2 [\sigma(x_i)P(x_i, t)]}{\partial x_i \partial x_j}$$

The quasi-potential $U = -\ln(P)$.

Numerical solution by **self-consistent mean field approximation** (Li & Wang, *PLoS Comput. Biol.* 2013).

Assumption: probabilities of gene concentrations are independent and follow Gaussian distribution

Gaussian process dynamical model (GPDM) is a probabilistic approach to model high dimensional time series data in a low dimensional latent space with a dynamical model.

In the GPDM framework, the original data consists of N points, $Y = [y_1 \dots y_N]$ with D dimensions are recovered from the latent space with the corresponding two dimensional data $X = [x_1 \dots x_N]$. The relationship between the latent space data and the high dimensional data with Gaussian noise added would be

$$y_i = Wx_i + \sigma_i$$

where $W \in R^{D*2}$ which has a prior as a spherical Gaussian distribution with zero mean and $\sigma_i \in R^{D*1}$ sampled from a spherical Gaussian distribution with 0 mean and $\beta^{-1}I$ covariance. Thus the likelihood for all data points is

$$p(Y|X, \beta) = \prod_{d=1}^D p(y_{:,d}|X, \beta) = \prod_{d=1}^D N(y_{:,d}|0, K)$$

The mapping between latent space and original space is through maximization of the above likelihood.

K is the kernel matrix, applied to guarantee the smoothness and nonlinearity of output function.

Reactions

Species	Reactions	Parameters	Initial
Name			
OCT4_Gene		0.0	
NANOG_Gene		0.0	
SOX2_Gene		0.0	
GATA6_Gene		0.0	
CDX2_Gene		0.0	
GCMF_Gene		0.0	
targetGene		0.01	
degradation		0.0	
p53		0.0	
A		25.0	
SG		0.0	
SN		0.0	
OCT4		0.01	

Set Parameters

Model: Deterministic (ODEs) | Stochastic (SDEs) | Blue Stochastic Simulation (jvarkit)

Number of measured points per time series: 1750

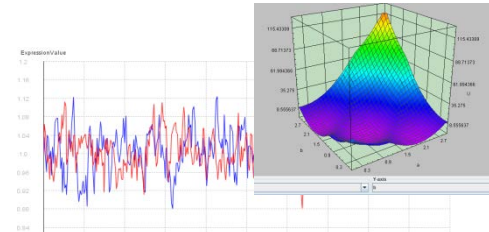
Number of measured points per time series: 257

Coefficient of noise term: 0.00

Output the benchmark (dataset + network files) will be saved: [Browse]

[Simulate] [Cancel]

- Solve ODE
- Solve SDE
- Stochastic model



LOAD

SAVE

TRAJ

LAND

- Load saved models
- Save simulation result
- Generate trajectories
- Construct Landscape

Control Panels

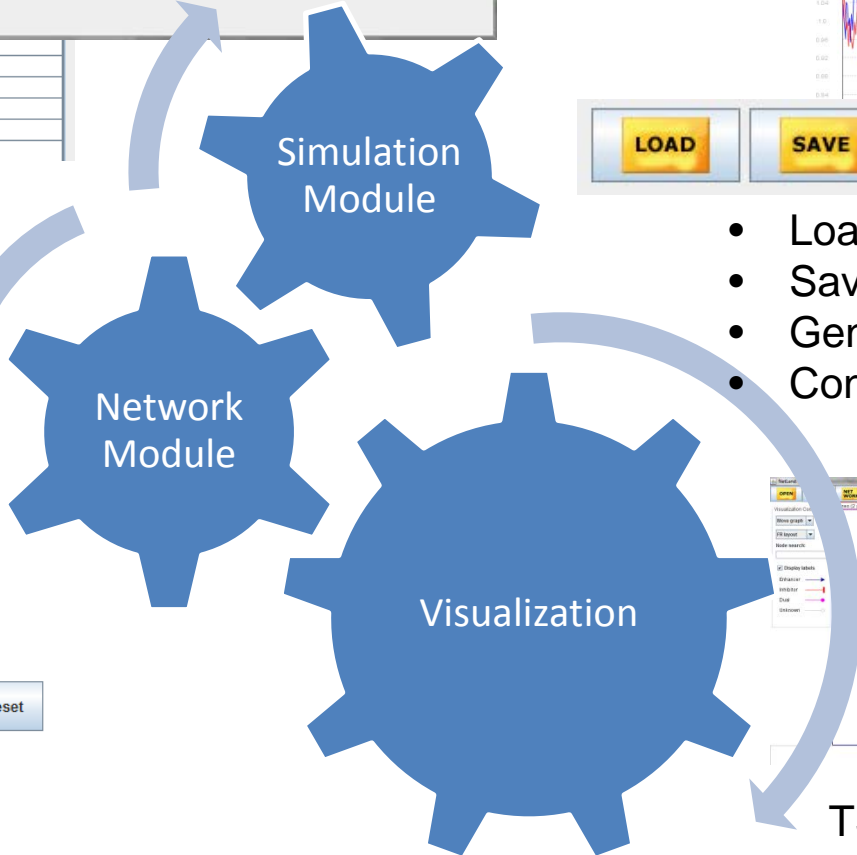
Screen the network

Block nodes

Add nodes and edges

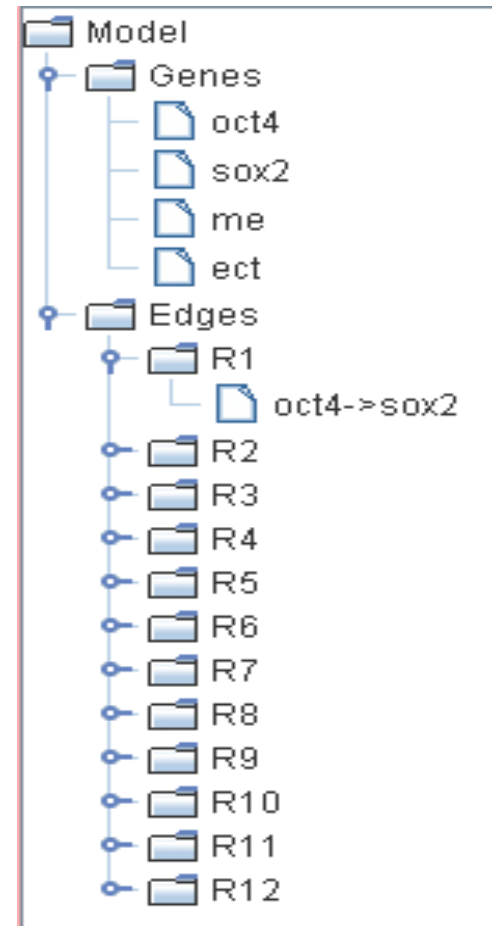
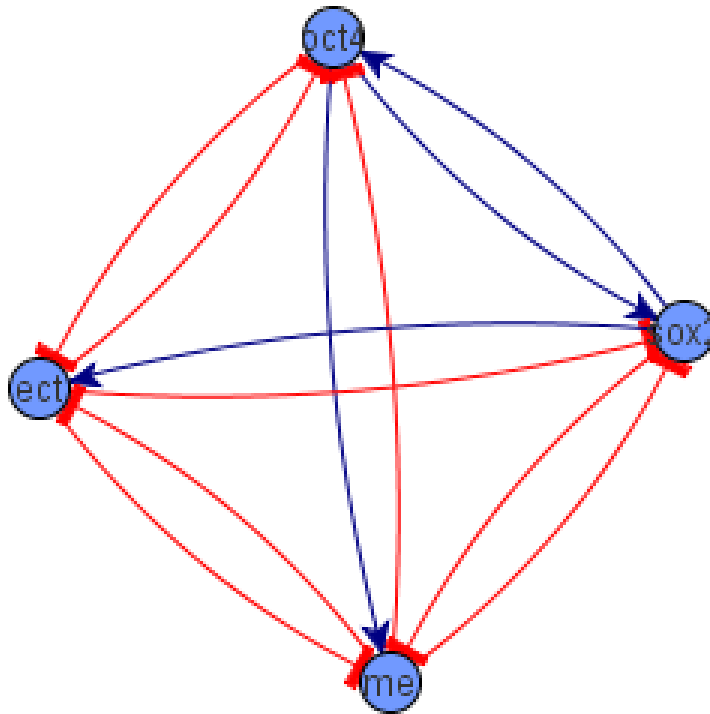
Reset

Modification of the network



TSV, GML, DOT, SBML

Example: A 4-gene network of stem cell fate



Species Panel

Species	Reactions	Parameters	Info
Name	InitialValue		
oct4	1.0		
sox2	1.0		
me	0.0		
ect	0.0		
Cell Volume	1000.0		

Reaction Panel

Species	Reactions	Parameters	Info
RxnId	TargetGene	Modifiers	Equation
R1	oct4	me,ect,oct4...	R1
R2	sox2	me,ect,oct4...	R2
R3	me	oct4,sox2,e...	R3
R4	ect	sox2,oct4,...	R4

Render

$$) + 1 * 0.7^s / (0.7^s + me^s) * 0.7^s / (0.7^s + ect^s) * (0.4 + oct4 * sox2 / (0.35 + oct4 * sox2)) -$$

Submit

Parameters Panel

Species	Reactions	Parameters	Info
Name	Value		
deg_oct4	1.0		
max_oct4	1.0		
BasalExpression_oct4	0.0		
k_oct4_0	0.5		
k_oct4_1	0.5		
k_oct4_2	0.5		
n_oct4_0	4.0		
n_oct4_1	4.0		
n_oct4_2	4.0		
l_oct4_0	1.0		
l_oct4_1	1.0		
l_oct4_2	1.0		
BasalExpression_oct4_0	0.0		
BasalExpression_oct4_1	0.0		
BasalExpression_oct4_2	0.0		
deg_sox2	1.0		
max_sox2	1.0		
BasalExpression_sox2	0.0		

Generation of trajectories

The 'Set Parameters' dialog box contains the following settings:

- Model: Deterministic (ODEs) (selected)
- Deterministic (ODEs): 10
- Stochastic (SDEs): 128
- Run Stochastic Simulation (max): 257
- Number of measured points per time series: 0.05
- Noise in the dynamics of the networks (SDEs): 0.05
- Coefficient of noise term: 0.05
- Output the benchmark (dataset + network files) will be saved: [Browse]

Buttons: Simulate, Cancel

- Solving ODEs
- Solving SDEs
- Stochastic model

Construction of the landscape

The 'Construction of the landscape' dialog box contains the following settings:

- Set Up Boundary for GeneExp: 3
- Set No of trajectories to generate: 100
- Set MaxTime for trajectories: 128
- Set Number of Points for each trajectory: 257
- Set Analysis Genes (seperated by ;): oct4;sox2;me;ect;
- Output landscape will be saved: [Browse]

Buttons: Probabilistic (selected), GPDM, Entropy, Run, Cancel

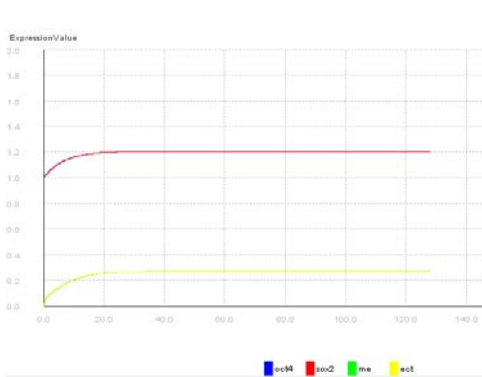
- Probabilistic Landscape
- Scaled Landscape

Deterministic model

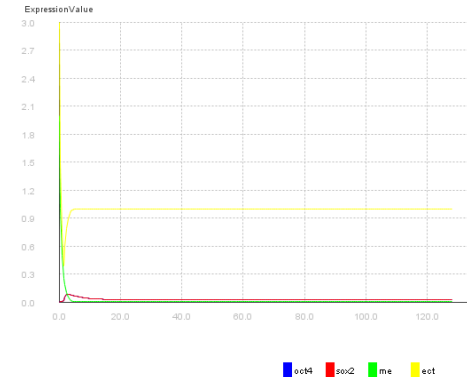
Runge-Kutta method to solve the **ODE** model

Euler-Maruyama method to get the approximate numerical solution of **SDEs**

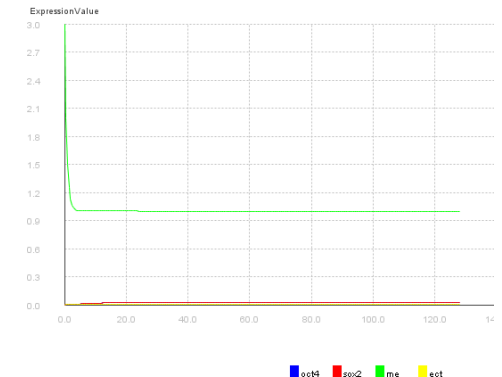
ODE



ESC

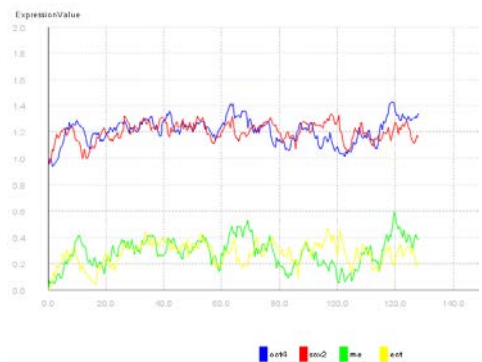


ECT

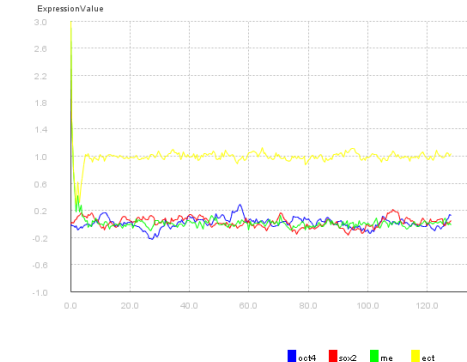


ME

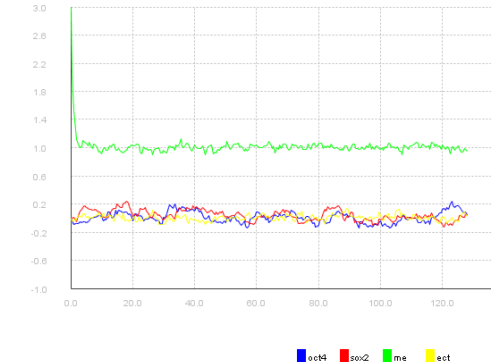
SDE



ESC



ECT

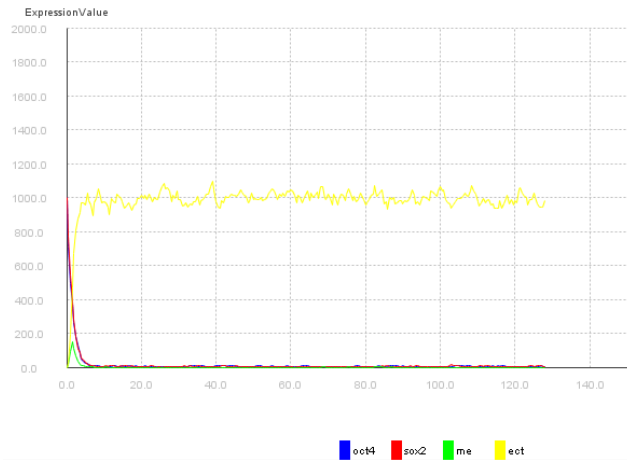


ME

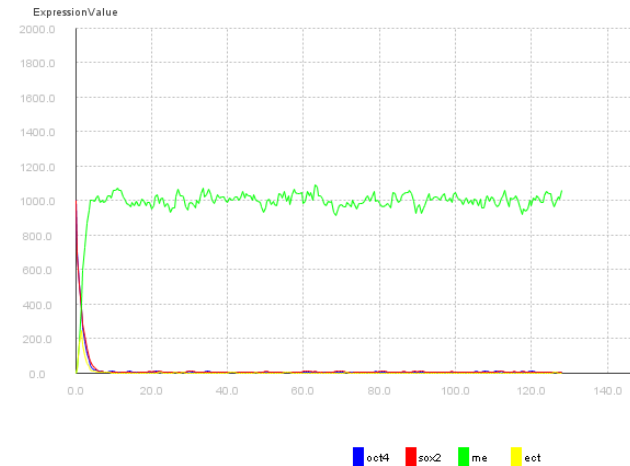
T=128, D=0.03

Stochastic model

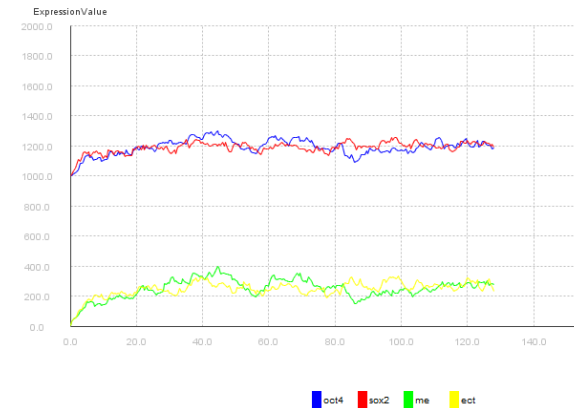
Gillespie algorithm is used for simulation



ECT

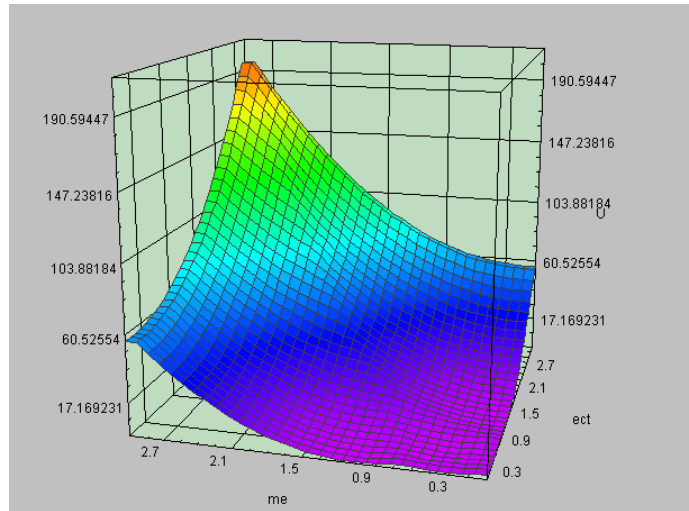


ME



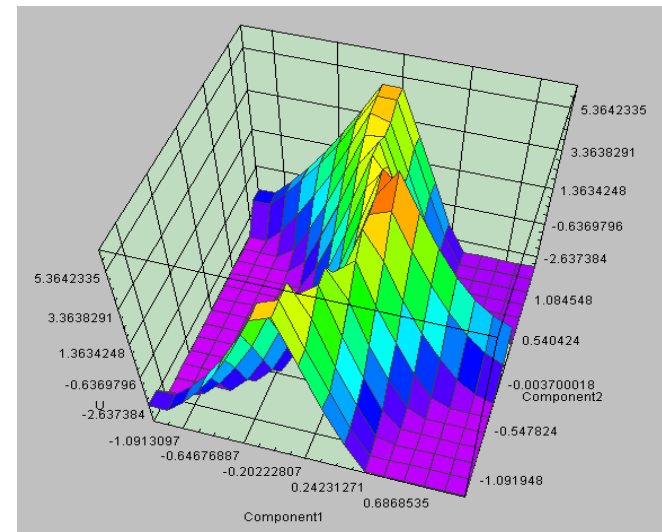
ESC

Visualization result



Visualization of global dynamics in a landscape. The high dimensional data is reduced to two dimensions as x-axis and y-axis.

It shows the landscape of the 10-gene network with expression values of Sox1 and Gata6 as x-axis and y-axis. Three attractors were located at (0,0), (2.5,0) and (0,2.5) representing stem cell state, ectoderm and mesoderm respectively.



Summary

- NetLand is a software tool for simulation and visualization of kinetics of transcriptional regulatory networks (GRN).
- Both deterministic (ODE) and stochastic (SDE) models are supported.
- Plot the Waddington's epigenetic landscape for GRNs of more than 3 genes.
- **Future work:**
 - Scalability of bigger networks (> 30 genes)
 - Validation with real data
 - Add discrete version (Boolean network)

Acknowledgments



Funding support

- MOE AcRF Tier 1 Seed Grant on Complexity (RGC 2/13)
- MOE AcRF Tier 2 Grant (ARC9/10), Ministry of Education, Singapore
- SCE-BII Joint PhD Program

Discussion with

Dr. Xiaomeng Zhang (Research Fellow)

Dr. Ket Hing Chong (Research Fellow)