行政院國家科學委員會專題研究計畫　成果報告

子計畫四：基於計算智慧之基因網路模型與穩定度之研究

(II)

計畫類別：整合型計畫
計畫編號：NSC94-2213-E-009-124-
執行期間：94 年 08 月 01 日至 95 年 07 月 31 日
執行單位：國立交通大學電機與控制工程學系(所)

計畫主持人：李祖添
共同主持人：吳幸珍
計畫參與人員：蔣欣翰，周家賢

報告類型：精簡報告
報告附件：出席國際會議研究心得報告及發表論文
處理方式：本計畫可公開查詢

中　華　民　國 95 年 9 月 28 日

行政院國家科學委員會補助專題研究計畫■ 成 果 報 告
□期中進度報告

智慧型系統在卵巢癌晶片之分子演化與控制-子計畫四：基於計算智慧之基因網路模型與穩定度之研究(II)

計畫類別：□ 個別型計畫　　■ 整合型計畫
計畫編號：NSC94－2213－E－009－124
執行期間：　　2005 年 8 月 1 日至 2006 年 7 月 31 日

計畫主持人：李祖添 國家講座
共同主持人：吳幸珍 助理教授
計畫參與人員： 蔣欣翰, 周家賢

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
□出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年■二年後可公開查詢

執行單位：交通大學電機與控制學系

中　華　民　國　　95 年　7　月　　31 日

智慧型系統在卵巢癌晶片之分子演化與控制
子計畫四：基於計算智慧之基因網路模型與穩定度之研究(II)

計畫編號: NSC 94-2213-E-009-126
執行期限:94/08/01 - 95/07/31

主持人：李祖添 講座教授　　共同主持人：吳幸珍 助理教授
參與人員：蔣欣翰, 周家賢
執行單位：國立交通大學電機與控制系

中文摘要

本年度計畫旨在以計算智慧方法來實現基因網路(pathway)之建模。一方面，由非線性微分方程式的基因網路建模，根據其生成項與消耗項之意義，來建立基因網路之 pathway；另一方面，利用 Bayesian Network，分別對卵巢癌基因和 Sacharomyces cerevisiae cell-cycle 基因時間點資料，做基因調控網路的建模。

關鍵詞：Bayesian Network，基因調控網路

英文摘要

In this project, several computational intelligent approaches are developed for constructing the gene-pathway to realize the interactions between genes. In other words, we shall achieve modeling and analysis of a gene network. First, the pathway is generated on the basis of nonlinear differential equations, which include mathematical descriptions of activatory terms and inhibitory terms. On the other hand, the gene regulatory networks are constructed by Bayesian Network technique. We have finish the pathways for two biology dataset (varian-cancer genes and Sacharomyces cerevisiae cell-cycle genes),

Keywords: Bayesian Network, Gene Regulatory Network

（一）目的與文獻

In this project, the gene-relationship pathways are proposed, respectively, by HDE/IGA-technology-based nonlinear differential equations and a Bayesian network.

The gene-relationship pathway for ovarian cancer can't be constructed by nonlinear-differential-equations-based approaches since we have only 3 time-courses data. In other words, the gene network for our ovarian-cancer dataset can't be obtained via identification of S-system [1] and lumped power-law system, modified by [2]. The research in cell cycle is important and the related dataset is available in the website. Therefore, we use yeast dataset to realize cell reproduction and further to predict cancer's development. In our work, HDE and IGA techniques are used to construct, respectively, S-system and lumped power-low system for Sacharomyces cerevisiae cell-cycle genes dataset, which is operated in cdc25 experiment. As we know, these two mathematical models are highly nonlinear. Therefore, the identification for both structure and parameters is tough work even for a system with low numbers of genes. However, the generated two systems are useful in realizing gene-genes relationship.

A genetic algorithm is an optimal searching method proposed by Holland in 1975 [3-9]. It is different from other optimal methods. No complexity calculation is needed in a genetic algorithm. However, the performance, at first, is not so good. Recently, researchers proposed various methods to increase the performance [10-14]. Chiang proposed a genetic algorithm with an improved-evolution-direction operator (IEDO) [15]. The IEDO is further evolved by evolution direction operator (Yamamoto [14]). More and more researches focused on the so-called improved genetic algorithm (IGA), where an IEDO, an elitism [5, 6], an acceleration operator [16-18] and a migration [19-21] are integrated into a genetic algorithm.

Evolution algorithms have been applied widely in parameter estimation. Storn and Price proposed a differential evolution (DE) method [22, 23]. In this method, population-differences are calculated during an evolution process, but the solution space frequently bog down into local minima. Chiou and Wang proposed a hybrid differential evolution (HDE) method to overcome this drawback [24]. A HDE can find the global optimal solution for a highly nonlinear differential system and can escape from bogging down into local optimal solution. Voit adopted this method, and further added the deviation for the slope of the estimated system from a true system to accelerate the searching speed and to adjust the parameters of estimated system [25-27]. Tominaga adopted a genetic algorithm with a structure skeletalizing algorithm to achieve the parameter estimation for a power-law differential equations [28, 29]. Kikuchi modified the crossover process and added the pruning terms into a fitness function [30]. In our project, we adopt both HDE and IGA algorithms to identify both S-system and lumped power-low system and to construct the corresponding gene network.

In other hand, we use a Bayesian Network to construct a gene network directly from the gene time-courses data. Bayesian Network is a graphical model, based on the joint probability distribution of random variables. Many research methods are based on this technology [31-36]. We here use this graphical technique to generate the gene networks for both ovarian-cancer and Sacharomyces cerevisiae cell-cycle genes dataset.

（二）研究方法與結果討論

2.1 Nonlinear D.E.-based Pathway

2.1.1 HDE-base S-system Gene Network

S-system use power-law flux to describe the synergism and saturation of the biological system [1],

$$\dot{X}_i = V_i^+ - V_i^- = \alpha_i \prod_{j=1}^{n} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n} X_j^{h_{ij}}, \quad \text{for } i = 1,2,...,n, \qquad (1)$$

where V_i^+ represents activatory term and V_i^- represents inhibitory term. X_i is the state variable or reactant; n is the number of X_i. α_i is the production rate-constant and β_i is the degradation rate-constant; both can be positive or zero. g_{ij} and h_{ij}, are kinetic orders; their values can be positive to indicate activating influences or negative to denote inhibition. The genes in V_i^+ affect activatory reaction of X_i. The genes in V_i^+ affect inhibitory reaction of X_i. We adopt HDE with modified collocation method and slope approximation method to obtain gene network in S-system form. The parameter estimation results are shown in Table 1.

Table 1. HDE-base S-system Gene Network

HDE								
	$\dot{X}_1$	$\dot{X}_2$	$\dot{X}_3$	$\dot{X}_4$	$\dot{X}_5$	$\dot{X}_6$	$\dot{X}_7$	$\dot{X}_8$
α	9.68	3.09	1.48	1.09	12.13	20.94	23.06	0.11
β	0.01	9.30	2.71	0.87	18.34	25.00	19.70	3.45
g_{i1}	-3.00		1	-3.00	-1.88	0.67	-1.44	
g_{i2}	1.02	2.73	-1.18	-3.00	0.45	-3.50	-3.45	
g_{i3}	-3.00		-2.00	-1.10	2.30	-2.56	-2.18	-3.28
g_{i4}	1.53	-1.85	-2.45		1.43		-3.91	
g_{i5}	-3.00	3.00	-3.73	0.80			-3.99	-3.08
g_{i6}	-2.08	0.58	-3.66		0.67	-2.15	-3.99	
g_{i7}		-3.00	-3.07			-3.50		
g_{i8}		-2.83	-1.08		0.53	-2.37		
g_{i9}		-1.08	3.11	3.00	1.34	3.15	-4.00	
g_{i10}			-1.10		0.48	-3.50		-2.25
g_{i11}		2.57	-3.51	0.77	-1.51	-2.58		-2.96
g_{i12}		2.46	-3.98	2.99	0.64	-3.50	3.32	-2.45
g_{i13}	0.56		-2.14	2.89	-2.40	-3.50	-2.89	-1.36
g_{i14}		2.57				-3.48		-2.34
g_{i15}		-2.50	-3.63		-0.59	-2.30	1.70	-0.74
g_{i16}	-2.99		-3.18		-1.22	-3.50	-3.84	
h_{i1}		1.38		1.71	0.60	2.96	-3.43	
h_{i2}	3.00		-0.52	-2.63	2.54	3.50	-0.92	-1.98
h_{i3}			-3.22	-1.32	-0.80	-3.50	2.12	-0.82
h_{i4}	-3.00		-2.59	3.00			1.41	1.78
h_{i5}	1.48	0.51	3.57					-1.21
h_{i6}	2.56		1.14	1.75	-1.48			4.00
h_{i7}								1.90
h_{i8}					-1.22			3.79
h_{i9}				-2.62	0.77	3.50		3.38
h_{i10}	3.00		3.82	2.44			1.33	-3.40
h_{i11}		-2.08		1.73	-1.93		1.57	-1.42
h_{i12}		2.09	-2.24	-2.60	0.86		3.54	-0.84
h_{i13}	-1.12		1.59		2.98	-3.08		-4.00
h_{i14}	2.99			-3.00	0.64			1.32
h_{i15}	3.00		2.78	-2.95	-2.92	-1.04		
h_{i16}				-3.00			-2.57	-3.55
	$\dot{X}_9$	$\dot{X}_{10}$	$\dot{X}_{11}$	$\dot{X}_{12}$	$\dot{X}_{13}$	$\dot{X}_{14}$	$\dot{X}_{15}$	$\dot{X}_{16}$
α	8.86	4.66	9.87	0.09	1.31	5.65	10.58	17.04
β	10.39	2.22	0.76	6.94	5.56	3.91	10.83	18.24
g_{i1}		-1.66		-3.97	-2.88	-2.46	-0.55	-3.96
g_{i2}	1.93	-0.91	-1.41			-3.99	-3.00	

g_{i3}	0.94	-1.90	-2.45	-4.00	2.05	-4.00	-2.98	-3.99
g_{i4}		-0.83	2.16	-3.81	-3.00	1.86	1.06	-2.89
g_{i5}	-1.53	-1.50	-2.58		-3.00	-2.38		-4.00
g_{i6}	2.71		-2.19		-3.00	0.88	-2.90	-4.00
g_{i7}		2.28		1.61	-2.67			
g_{i8}	-0.63		-2.40		-2.26			
g_{i9}		0.55		-2.86		-2.97		2.80
g_{i10}		-1.56	-2.91		-0.86	1.94		
g_{i11}	-0.97		-2.96	3.88	-3.00	-4.00		-3.99
g_{i12}	1.59	1.73	-1.95		-1.54			-3.93
g_{i13}		1.35	-2.99	-1.32				-1.74
g_{i14}	-0.68	1.94	1.38	0.67	-1.62	-3.55		
g_{i15}	-1.53	2.93	-1.68				-1.50	-0.85
g_{i16}	-0.73	-1.49	-1.05	0.61	1.95			-4.00
h_{i1}		-0.83	-1.08			1.23	1.32	
h_{i2}	-2.28	1.61		-2.65	1.94	2.65	-2.46	
h_{i3}	1.65	1.26		-3.16	1.27			1.86
h_{i4}	1.10	2.07	1.75	-2.11	1.44	2.80		-0.81
h_{i5}		-1.93			1.75		0.89	
h_{i6}		2.20	-2.13	3.56	-2.97	-4.00	-2.71	
h_{i7}		1.56	-1.97		-2.99		-1.40	
h_{i8}	-1.05		0.78	3.98			3.00	
h_{i9}	1.09		3.00		1.92		2.13	4.00
h_{i10}	1.78	-2.85		2.48	-2.33		3.00	
h_{i11}		1.46		-2.99	0.52			
h_{i12}	-0.51		1.48	4.00	1.91		1.44	
h_{i13}	-1.80	1.26	-0.69	-3.99	3.00			
h_{i14}			-3.00	2.37			-2.91	2.90
h_{i15}				3.51			-3.00	
h_{i16}	0.76	-2.51		-0.92	0.58		2.70	-3.97

Five sets of time-course data are generated from different initial values. Variable X_1 denotes CDC28, X_2 for CLN3, X_3 for SWI4, X_4 for SWI6, X_5 for MBP1, X_6 for FUS3, X_7 for FAR1, X_8 for CLN1, X_9 for CLN2, X_{10} for SIC1, X_{11} for CLB5, X_{12} for CLB6, X_{13} for CDC6, X_{14} for CDC20, X_{15} for GRR1 and X_6 for CDC4. After training, we obtain the S-system structure. Based on this structure, we can create the gene regulatory network in Fig. 1. Black bold lines, growing themselves, represent synthetic influence; black bold lines, growing outside, represent degraded influence. Black lines represent activation reaction and red lines represent inhibition reaction. The start point of the lines is the reactant and the end point is the product. Biochemical reactions can be described by the reconstructed model. We now can realize the interaction between various genes from the gene regulatory network. The increasing in concentrations of CLN3, MBP1, FUS3, CLB5, CLB6 and CDC20 will bring increasing synthetic influence in concentration of CLN3. However, the increasing in concentrations of SIW6, FAR1, CLN1 and CLN2 will bring decreasing synthetic influence in concentration of CLN3.

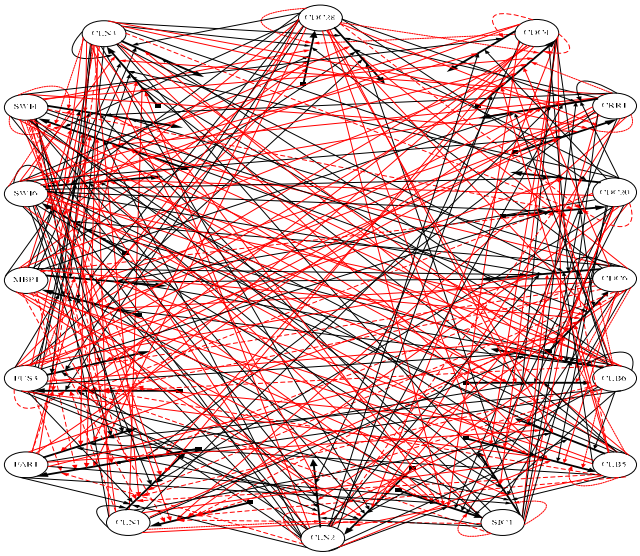


Figure 1. S-system Gene Network (JD pathway)

2.1.2 IGA-based lumped power-law gene network

The second mathematical model is approximated from the model adopted in [6, 7].

$$\dot{X}_i(t) = G_i(t) - \lambda_i X_i(t), \quad i=1,2,\cdots,n, \qquad (2)$$

where $G_i(t)$ is the transcription rate, λ_i is the self-degradation rate and n is the number of the variable, $X_i(t)$ is the concentration of the i-th gene at time t. $G_i(t)$ is a nonlinear function,

$$G_i(t) = \sum_{j=1}^{m} a_{ij} \frac{1}{1+\exp\left\{-\alpha\left[u\left(t,\beta_j,\delta\right)-\gamma\right]\right\}}, \qquad (3)$$

$$u(t,\beta_j,\delta) = \begin{cases} 0 & t \le \beta_j - \delta \\ \dfrac{t-\left(\beta_j-\delta\right)}{\delta} & \beta_j - \delta \le t \le \beta_j \\ \dfrac{(\beta_j+\delta)-t}{\delta} & \beta_j \le t \le \beta_j + \delta \\ 0 & t \ge \beta_j + \delta \end{cases}. \qquad (4)$$

We use a power-law function $V_i(t)$ to approximate the nonlinear function $G_i(t)$ in Eq. (3) to denote the synthesis rate.

$$V_i(t) = \lambda_i \prod_{j=1}^{n} X_j^{f_{ij}}(t), \qquad (5)$$

where λ_i is the rate constant and f_{ij} is kinetic order. Further, the degradation term in Eq. (2) is replaced by $\gamma_i X_i^{k_i}(t)$ to emphasize how a gene reacts itself. The kinetic orders, f_{ij} and γ_i, can be positive or negative; positive kinetic orders indicate activating influences, but negative kinetic orders mean inhibition. In other words, the following lumped power-law system is proposed.

$$\begin{aligned} \dot{X}_i(t) &= f_i(\mathbf{X},\mathbf{P}) = V_i(t) - \gamma_i X_i^{k_i}(t) \\ &= \lambda_i \prod_{j=1}^{n} X_j^{f_{ij}}(t) - \gamma_i X_i^{k_i}(t), \quad i=1,2,\cdots,n, \end{aligned} \qquad (6)$$

where n is the number of the variables; the vector **X** in Eq. (6) indicates all genes in the yeast cell cycle; the vector **P** in Eq. (6) consists the rate constants, λ_i and γ_i, and kinetic orders, f_{ij} and k_i. We adopt HDE with modified collocation method and slope approximation method to obtain gene network in lumped power-law form. The parameter estimation results are shown in Table 2.

Table 2. IGA-based lumped power-law gene network

IGA								
	$\dot{X}_1$	$\dot{X}_2$	$\dot{X}_3$	$\dot{X}_4$	$\dot{X}_5$	$\dot{X}_6$	$\dot{X}_7$	$\dot{X}_8$
γ	13.46	13.27	8.07	11.06	0.41	0.01	6.23	11.35
λ	14.40	14.44	2.54	4.29	18.33	8.25	11.07	4.98
f_{i1}	3.74	-4.00	1.41	-1.38	0.87	3.10	0.92	
f_{i2}	-0.49	2.78	2.68	-2.24	0.56		1.82	0.82
f_{i3}	-4.00	1.03	-2.51	0.77		2.27	-4.00	-1.66
f_{i4}				4.00			-1.93	-4.00
f_{i5}	4.00	-3.13	-2.54	2.36	3.97	1.21		-0.54
f_{i6}	-4.00	-2.20	-1.13		2.47			1.08
f_{i7}	-3.85				-3.53	-1.83	1.43	0.07
f_{i8}	4.00		-3.99	-2.33	1.14	3.49	1.30	-1.91
f_{i9}		-4.00			-3.75	-1.75	-2.27	-1.46
f_{i10}			2.23		-0.67	-4.00		
f_{i11}	-3.50		1.60			-1.52	-1.59	0.93
f_{i12}	-2.04		0.60	2.01	2.97		2.54	3.69
f_{i13}	3.33	2.83	-1.59	3.06	-3.97	-2.45		3.49
f_{i14}	1.26	2.36	1.38	-3.96	-3.84	1.68	3.24	-0.91
f_{i15}	4.00	3.85	0.83			-2.98	-0.92	0.78

f_{i16}	-4.00	1.35		-1.94	2.40	3.80	-0.64	
k_i	-3.90	-3.86	-1.35	-2.73	-3.80	-0.34	-1.93	-3.34

	$\dot{X}_9$	$\dot{X}_{10}$	$\dot{X}_{11}$	$\dot{X}_{12}$	$\dot{X}_{13}$	$\dot{X}_{14}$	$\dot{X}_{15}$	$\dot{X}_{16}$
γ	20.00	2.48	7.42	6.37	8.26	0.27	14.94	11.41
λ	5.06	13.53	2.56	9.92	2.99	0.24	12.15	17.42
f_{i1}	-0.62	-2.68		1.39	2.19	-1.15	-2.46	
f_{i2}	-1.36	4.00	3.85	0.50	0.60		-1.92	2.67
f_{i3}	1.66	-2.30	3.93	3.34	-1.95	-2.21	-1.92	-2.78
f_{i4}				-2.94	-2.58	1.32	1.76	0.90
f_{i5}	2.12	-2.21	-3.80	-2.23	1.61	1.16	0.68	1.90
f_{i6}			-3.19	-0.84	1.14	2.68	1.91	-3.72
f_{i7}	1.76	3.63	-2.39		-1.43	-0.59	-1.22	
f_{i8}	-3.75		-4.00	-2.44	-0.88	1.92	2.23	-0.57
f_{i9}	-2.91	-1.36	1.46	-0.94			1.30	
f_{i10}					3.41		-0.68	
f_{i11}	0.72	-3.80		-2.88	2.23	-2.81	-0.92	1.75
f_{i12}	0.62			2.63	-1.82	-1.43		-0.63
f_{i13}	0.51		4.00	4.00		0.78		3.36
f_{i14}	-1.08	1.26	0.78		0.96	0.78	-4.00	-1.92
f_{i15}	4.00	2.11	0.58	-0.72	-3.75	-1.57		-3.96
f_{i16}	-1.20			4.00		0.86	2.94	0.82
k_i	-0.49	-1.68	-1.49	-3.54		0.52	-3.90	-0.51

Fig. 2 is the pathway for the modified power-low model. Black lines represent activation reaction and red lines represent inhibition reaction. The start point of the lines is the reactant and the end point is the product. For instance, the concentration of CDC28 increases rapidly as the concentrations of MBP1, CLN1, CDC6, CDC20 and GRR1 increase; however, the concentration of CDC28 decreases rapidly as the concentrations of CLN3, SWI4, FUS3, FAR1, CDC4, CLB6 andCLB6 increase.

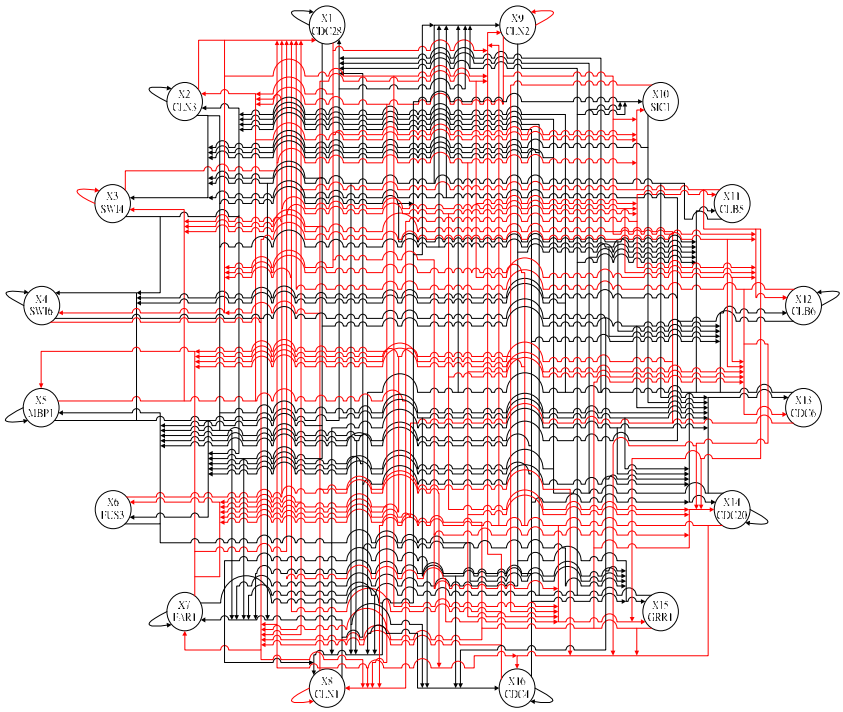


Figure 2. Lumped Power-law Gene Network (JC pathway)

2.2 Bayesian-method-based Gene Network

2.2.1 Bayesian Network

A Bayesian Network is a graphical model that encodes the joint probability distribution for a set of random variables. Assume $D = (V, E)$ be Directed Acyclic Graph (DAG), where *V* represents nodes and *E* represents directed edges (arrows) between the nodes. The DAG is the structure of the Bayesian Network. Every node $v \in V$ has a corresponding random variable X_v. The set of random variables defines $X = (X_v)_{v \in V}$. Each node with its parents denoted $pa(v)$ has a local probability distribution $p(x_v \mid x_{pa(v)})$. The start point of directed edge *E* denotes $pa(v)$ and end point denotes node v. The Bayesian Network for a set of random variable X defines the pair (D, P), where P is a set of local probability distributions for all random variables.

Assume conditional independencies between the random variables *X* are established, the directed edges in *D* can represent relationship for the random variables *X* through the factorization of the joint probability distribution,

$$p(x) = \prod_{v \in V} p(x_v \mid x_{pa(v)}) , \tag{1}$$

The flowchart for Bayesian Network is shown Figure 3. After gene time-courses data import, master prior procedure is proceeded. First, the local probability distributions $p(x_v \mid x_{pa(v)})$ for all random variables are calculated. Assume all local probability distributions are Gaussian linear regressions, We parameterize this as $\theta_v = (m_v, \beta_v, \sigma_v^2)$ so that

$$(X_v \mid x_{pa(v)}, \theta_v) \sim N(m_v + \beta_v x_{pa(v)}, \sigma_v^2), \tag{2}$$

where m_v is regression intercept, β_v is regression coefficients, σ_v^2 is conditional variance. Thus for each configuration of v, the distribution of X_v is Gaussian with mean and variance given as in equation (2). Then, the joint probability distribution for X is a conditional Gaussian distribution with density of the form

$$p(x \mid \theta) = \left|2\pi\Sigma_v\right|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(x - M_v)^T \Sigma_v^{-1} (x - M_v)\right\}, \tag{3}$$

where M_v is unconditional mean and Σ_v is covariance matrix. From the joint probability distribution, the marginal distribution of all parameters can be determined. Therefore, we define this the master prior procedure.

Next step is parameter learning. To estimate the parameters in the network, we use the Bayesian approach. We use our uncertainty about parameters θ in a prior distribution $p(\theta)$, use data *d* to update this distribution, and hereby obtain the posterior distribution $p(\theta \mid d)$ by using Bayes' Theorem,

$$p(\theta \mid d) = \frac{p(d \mid \theta) p(\theta)}{p(d)}, \qquad \theta \in \Theta \tag{4}$$

where

$$p(d \mid \theta) = \prod_{x^c \in d} p(x^c \mid \theta). \tag{5}$$

Here $p(d \mid \theta)$ is the joint probability distribution of *d*, and called the likelihood of θ. *d* is a random sample from the probability distribution $p(x \mid \theta)$. Θ is parameter space. For fixed *d*, $p(d)$ can be considered as a normalizing constant. Therefore, equation (4) can be expressed as

$$p(\theta \mid d) \propto p(d \mid \theta) p(\theta). \tag{6}$$

The purpose of parameter learning is to expect to find posterior parameter. Assume there is independence between θ, then

$$p(\theta \mid d) = \prod_{v \in V} p(\theta_{v \mid x_{pa(v)}} \mid d). \tag{7}$$

That is，posterior parameter is independent.

Posterior parameter θ with independence is obtained by parameter learning. Then, we can have the initial network and presume it is optimal temporarily. We refer to network score as following form,

$$p(D, d) = p(d \mid D) p(D), \tag{8}$$

where $p(d \mid D)$ is the likelihood of *D* and $p(D)$ is prior probability. We choose to let all DAGs be equally likely, then

$$p(D, d) \propto p(d \mid D). \tag{9}$$

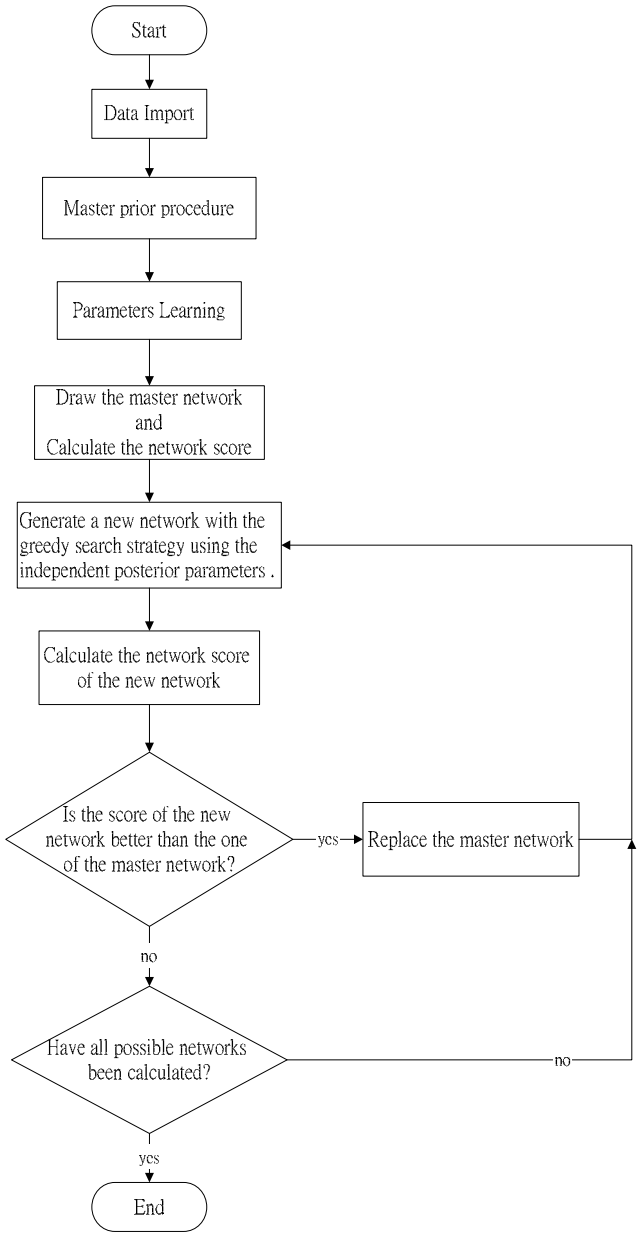


Figure 3. Bayesian Network Flow Chart

$$
\begin{aligned}
p(d \mid D) &= \int_{\theta \in \Theta} P(d \mid \theta, D) P(\theta \mid D) \\
&= \prod_{v \in V} \int \prod_{c: x^c_{pa(v)} = x_{pa(v)}} p(x^c_v \mid x^c_{pa(v)}, \theta_{x|x_{pa(v)}}, D)\, p(\theta_{x|x_{pa(v)}} \mid D) d\theta_{x|x_{pa(v)}} \\
&= \prod_{v \in V} \frac{X((\rho_{x|x_{pa(v)}} + |b|)/2)}{X(\rho_{x|x_{pa(v)}}/2)\left[\det(\rho_{x|x_{pa(v)}} s_{x|x_{pa(v)}} \pi)\right]^{\frac{1}{2}}} \times \\
&\quad \left[1 + \frac{1}{\rho_{x|x_{pa(v)}}} (x^b_v - z^b_{pa(v)} \mu_{x|x_{pa(v)}}) s^{-1}_{x|x_{pa(v)}} (x^b_v - z^b_{pa(v)} \mu_{x|x_{pa(v)}})^T\right]^{\frac{-(\rho_{x|x_{pa(v)}} + |b|)}{2}}.
\end{aligned} \tag{10}
$$

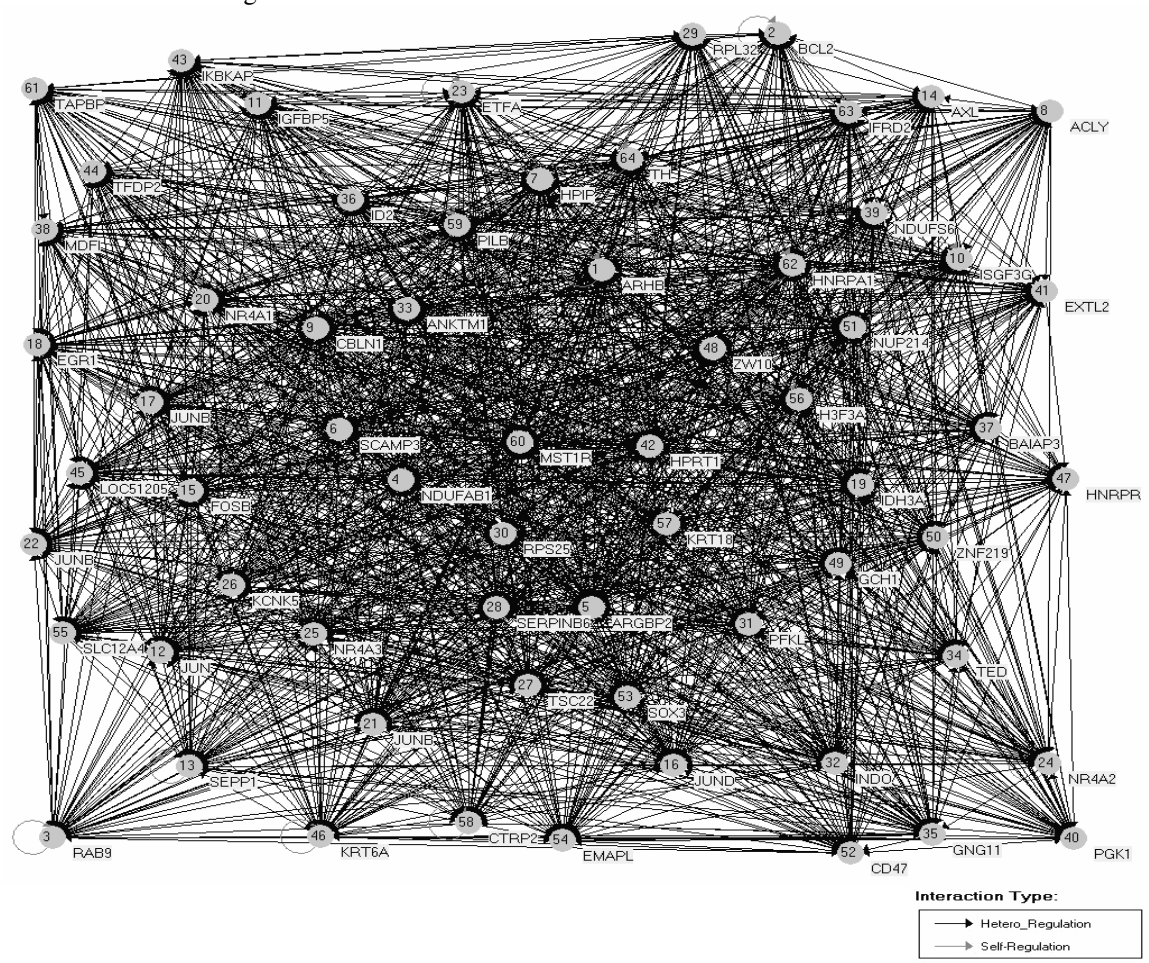
Then, the network score can be obtained simply. The final step is optimization of network. A strategy for searching for DAGs with high score is needed. We adopt the search strategy greedy search with random restarts. In greedy search, we create model D^* from model D that differ only by a single arrow, either added, removed or reversed and then calculate the network score. We compare two different models by posterior odds,

$$
\frac{p(D \mid d)}{p(D^* \mid d)} = \frac{p(D, d)}{p(D^*, d)} = \frac{p(D)}{p(D^*)} \times \frac{p(d \mid D)}{p(d \mid D^*)}, \tag{11}
$$

where $p(D)/p(D^*)$ is prior odds and $p(d|D)/p(d|D^*)$ is Bayes Factor. Because D and D^* are different by a direction point, the prior odds approximate to 1. Therefore, if Bayes Factor is greater than 1, D is better than D^*. In contrast, if Bayes Factor is smaller than 1, D^* is better than D. After the procedure, the optimal network is obtained until we can't find any network is better than it.

2.2.2 Ovarian Cancer data set

Ovarian cancer data set which includes 64 genes is supported form subproject I. The gene regulatory obtained by Bayesian Network is shown in figure 4.



2.2.3 Sacharomyces cerevisiae cell-cycle data set

The gene time-courses data of Sacharomyces cerevisiae obtained by world wide web is used to construct gene regulatory network by Bayesian Network, which includes 16 genes. Here, we concern two kinds of data set, one is raw data and the other is smooth data generated by Matlab. The network score of raw data is -1708.558 as shown in figure 5.

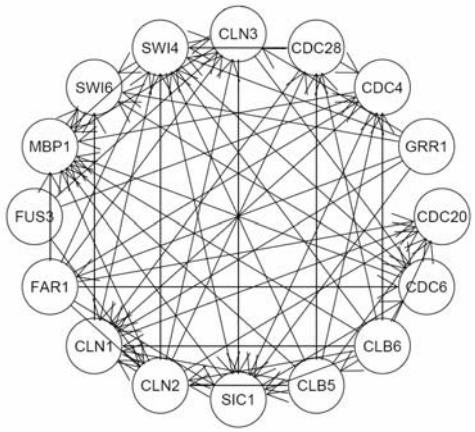


Figure 5. Gene regulatory network of raw data

The relationship of each node and its parents:

- [CDC28|CLN2:CLB5:CLB6]
- [CLN3|SWI4:FUS3:FAR1:CLN2:GRR1]
- [SWI4|CDC28:FAR1:CLN2:CLB5:CLB6:GRR1]
- [SWI6|SWI4:MBP1:CLN1:CDC6:GRR1]
- [MBP1|CLN3:SWI4:FAR1:CLN2:SIC1:CLB5:CLB6:CDC6:CDC4]
- [FUS3]
- [FAR1|GRR1]
- [CLN1|CDC28:SWI4:CLN2:CLB5:CLB6:GRR1:CDC4]
- [CLN2|FUS3:FAR1:CLB5:CLB6]
- [SIC1|CLN3:SWI4:FAR1:CLB5:CLB6:GRR1:CDC4]
- [CLB5|CLB6]
- [CLB6]
- [CDC6|CLN3:SWI4:FAR1:SIC1]
- [CDC20|FAR1:CLN1:CLN2:CLB5:CLB6]
- [GRR1]
- [CDC4|CDC28:FUS3:FAR1:CLB5:CLB6]

The network score of smooth data is -28978.00 as shown in figure 6.

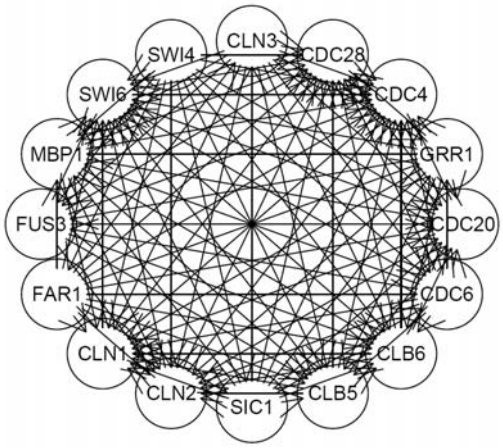


Figure 6. Gene regulatory network of smooth data

The relationship of each node and its parents:

- [CDC28|CLN3:SWI4:MBP1:FUS3:FAR1:CLN1:CLN2:SIC1:CLB5:CLB6: CDC6:CDC20:GRR1]
- [CLN3|FAR1:SIC1]
- [SWI4]
- [SWI6|CDC28:CLN3:SWI4:MBP1:FUS3:FAR1:CLN1:CLN2:SIC1:CLB5: CLB6:CDC6:CDC20:GRR1]
- [MBP1|CLN3:SWI4:FAR1:SIC1:CDC6:GRR1]
- [FUS3|CLN3:SWI4:CDC6:GRR1]
- [FAR1|SWI4]
- [CLN1|CLN3:SWI4:MBP1:FAR1:CLN2:SIC1:CLB5:CLB6:CDC6:CDC20: GRR1]
- [CLN2|CLN3:SWI4:MBP1:FUS3:FAR1:SIC1:CLB5:CLB6:CDC6:GRR1]
- [SIC1|SWI4:FAR1]
- [CLB5|CLN3:SWI4:MBP1:FUS3:FAR1:SIC1:CDC6:GRR1]
- [CLB6|CLN3:SWI4:MBP1:FUS3:FAR1:SIC1:CLB5:CDC6:GRR1]
- [CDC6|CLN3:SWI4:FAR1:SIC1]
- [CDC20|CLN3:SWI4:MBP1:FUS3:FAR1:CLN2:SIC1:CLB5:CLB6:CDC6: GRR1]
- [GRR1|CLN3:FAR1:SIC1:CDC6]
- [CDC4|CDC28:CLN3:SWI4:SWI6:MBP1:FUS3:FAR1:CLN1:CLN2:SIC1: CLB5:CLB6:CDC6:CDC20:GRR1]

（三）參考文獻

[1] E.O. Voit, *Computational Analysis of Biochemical System*, Cambridge University Press, 2000.

[2] H.C. Chen, H.C. Lee, T.Y. Lin, W.H. Li and B.S. Chen, "Quantitative characterization of the transcriptional regulatory network in the yeast cell cycle", *Bioinformatics*, Vol.20, pp. 1914-1927, 2004.

[3] J. H. Holland, "Genetic algorithm and the optimal allocation of trials", *SIMA J. Computing*, vol.2, no.2, pp.88-105, 1973.

[4] J. H. Holland, *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbor, 1975.

[5] T. Back, D. B. Fogel, and Z. Michalewicz, *Handbook of evolutionary computation*, University Oxford Press,

New York, 1996.
[6] Z. Michalewicz, *Genetic algorithms + data structures = evolution programs*, Springer-Verlag Berlin Heidelberg, New York, 1996.
[7] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley, Reading, MA, 1989.
[8] L. Davis, *Handbook of genetic algorithms*, Van Nostrand Reinhold, New York, 1991.
[9] V. S. Summanwar, V. K. Jayaraman, B. D. Kulkarni, H.S. Kusumakar, K. Gupta, and J. Rajesh, "Solution of constrained optimization problems by multi-objective genetic algorithm", *Computers and Chemical Engineering*, vol.26, pp.1481-1492, 2002.
[10] Y. H. Song and C. S. V. Chou, "Advanced engineer-conditioning genetic approach to power economic dispatch", *IEE Proc. – Gener. Transm. Distrib.*, vol.144, no.3, pp.285-293, 1997.
[11] S. Forrest, "Genetic algorithms: principles of natural selection applied to computation", *Science*, vol.261, pp.872-878, 1993.
[12] T. Back, *Evolutionary algorithms in theory and practice*, Oxford University Press, New York, 1995.
[13] L. Davis, *Genetic algorithms and simulated annealing*, Morgan Kaufmann Publishers, San Mateo, CA, 1987.
[14] K. Yamamoto and O. Inoue, "New evolutionary direction operator for genetic algorithms", *AIAA Journal Technical Notes*, vol.33, no.10, pp.1990-1993, 1995.
[15] Chao-Lung Chiang, Ching-Tzong Su, and Feng-Sheng Wang, "Augmented Lagrangian Method for Evolutionary Optimization of Mixed-Integer Nonlinear Constrained Problems", *Intern. Math Journal*, Vol. 2, no. 2, 119-154, 2002.
[16] R. Hooke and T. A. Jeeves, "Direct search solution of numerical and statistical problems", *Journal of the ACM*, vol.8, No.2, pp.212-229, 1961.
[17] W. Spendley, G. R. Hext, and F. R. Himsworth, "Sequential application of simplex designs in optimization and evolutionary operation", *Technometrics*, vol.4, p.441, 1962.
[18] Singiresu S. Rao, *Engineering optimization –theory and practice*, the 3rd ed., chapter 6, John Wiley & Sons , 1996.
[19] J. P. Chiou and F. S. Wang, "A hybrid method of differential evolutionary with application to optimal control problems of a bioprocess system", *in Proceedings 1998 IEEE on Evolutionary Computation Conference*, pp.627-632.
[20] Y. C. Lin, F. S. Wang, and K. S. Hwang, "A hybrid method of evolutionary algorithms for mixed-integer nonlinear optimization problems", *in Proceedings 1999 IEEE on Evolutionary Computation Conference*, pp.2159-2166.
[21] Y. C. Lin, K. S. Hwang, and F. S. Wang, "Plant scheduling and planning using mixed-integer hybrid differential evolutionary with multiplier updating", *in Proceedings 2000 IEEE on Evolutionary Computation Conference*, pp.593-600.
[22] Storn, R., and Price, K. V., "Minimizing the Real Function of the ICEC'96 Contest by Differential Evolution", *IEEE Conf. On Evolutionary Computation*, Nagoya, 842-844, 1996.
[23] Storn, R., and Price, K. V., "Differential Evolution: A simple and Efficient Heuristic for Global Optimization Over Continuous Spaces", *J. Glob. Optim.*, 11, 341-359, 1997.
[24] Wang, F.S., Su, T.L., and Jang, H.J., "Hybrid Differential Evolution for Problems of Kinetic Parameter Estimation and Dynamic Optimization of an Ethanol Fermentation Process", *Ind. Eng. Chem. Res.*, 40, 2876-2885, 2001.
[25] Savageau, M.A., and Voit, E.O., "Power-Law Approach to Modeling Biological Systems; II. Application to Modeling Biological Systems", *J. Ferment. Technol.*, 60, 229-232, 1982.
[26] Almeida, J.S., and Voit, E.O., "Neural-Network-Based Parameter Estimation in S-System Models of Biological Networks", *Genome Informatics*, 14, 114-123, 2003.
[27] Voit, E.O., and Almeida, J., "Decoupling Dynamics Systems for Pathway Identification from Metabolic Profiles", *Bioinformatics*, in press, 2004.
[28] Tominaga, D., and Okamoto, M., "Design of canonical model complex nonlinear dynamics", *In Proceedings of the International Conference on Computer Applications in Biotechnology*, 85-90, 1998.
[29] Tominaga, D., Koga, N. and Okamoto, M., "Efficient numerical optimization algorithm based on genetic algorithm for inverse problem", *In Proceeding of the Genetic and Evolutionary Computation Conference*, 251-258, 2000.
[30] Kikuchi, S., Tominaga, D., Arita, M. Takahashi, K., and Tomita, M., "Dynamic modeling of genetic networks using genetic algorithm and S-system", *Bioinformatics*, 19, 643-650, 2003.
[31] Husmeier, D., "Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks", *Bioinformatics*, vol. 19, no. 17, p.p. 2271-2282, 2003.
[32] Imoto S, Higuchi T, Goto T, Tashiro K, Kuhara S, Miyano S., "Combining microarrays and biological knowledge for estimating gene networks via bayesian networks", *J. Bioinform. Comput. Biol.*, vol. 2, no. 1, p.p. 77-98, 2004.
[33] Murphy, K., Mian, S., "Modelling Gene Expression Data using Dynamic Bayesian Networks", *Technical report*, Computer Science Division, University of California, Berkeley, CA, 1999.

[34] Bøttcher, S.G., "Learning Bayesian networks with mixed variables", *In Proceedings of the Eighth International Workshop in Artificial Intelligence and Statistics*, 2001.
[35] Bøttcher, S. G., Dethlefsen, C., "Learning Bayesian Networks with R", *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, March 20-22, Vienna, Austria.
[36] Bøttcher, S. G., Dethlefsen, C., "deal: A Package for Learning Bayesian Networks", *Journal of Statistical Software*, vol. 8, no. 20, p.p. 1-40, 2003.

(四)計畫成果自評

The gene regulatory networks/pathways are constructed, respectively, by nonlinear–mathematical-system-based methods and a Bayesian Network. We adopt two computational intelligent approaches (IGA and HDE methods). Two kinds of gene network are constructed via proposed S-system and lumped power-law systems, respectively. On the other hand, we use a stochastic and graphical approach, a Bayesian network, to construct the regulatory networks of ovarian-cancer and Sacharomyces-cerevisiae genes, respectively. The results has been accepted for publication in IEEE-SMC'06 and IEEE-EMBC'06 conference.

Publications:

1. Wu, S. J., Wu, C.T., Chou, C.H. and Lee, T.T., "Evolution-Based Gene Regulatory Network of Yeast Cell Cycle", *IEEE International Conference on Systems, Man, and Cybernetics*, 2006.
2. Wu, S. J., Chou, C.H., Wu, C.T. and Lee, T.T., "Inference of Genetic Network of Xenopus Frog Egg: Improved Genetic Algorithm", *IEEE International Conference of the Engineering in Medicine and Biology Society*, 2006.
3. Wu, S. J., Wu, C.T. and Lee, T.T., "Computation Intelligent for Eukaryotic Cell-Cycle Gene Network", *IEEE International Conference of the Engineering in Medicine and Biology Society*, 2006.