

## Knock-out 데이터를 이용한 유전자 조절망의 구성

홍성룡\*, 손기락\*

### Constructing Gene Regulatory Networks using Knock-out Data

Sung-Ryong Hong\*, Kirack Sohn\*

#### 요약

유전자 조절망은 유전자의 발현이 다른 유전자에게 영향을 주는 것을 표현하는 유전자 망이다. 오늘날 마이크로어레이 실험으로부터 유전자의 발현량을 측정된 대용량의 데이터가 이용 가능하다. 전형적인 데이터중의 하나는 특정 유전자를 제거한 후 다른 유전자의 발현량을 측정된 steady-state data이다. 본 논문은 이런 측정 데이터를 이용하여 중복 정보를 최소화하는 유전자 조절망을 재구성하는 방법을 제시한다. 제시한 모델은 기존 연구에서는 고려되지 않았던 사이클 형태로 나타나는 자동 조절 기능을 고려하였고, 또한 유전자의 억제자 또는 촉진자 역할을 고려하였다.

#### Abstract

A gene regulatory network is a network of genes representing how genes influence the activities of other genes. Nowadays from microarray experiments, a large number of measurements on the expression levels of genes are available. One of typical data is the so-called "steady-state model" data measuring the expression levels of other genes after knocking out a particular gene. This paper shows how to reverse engineer a parsimonious gene regulatory network, using these measurement data. Our model considers auto-regulation, which forms a cycle in a genetic network. We also model repressor and enhancer roles of genes, which are not considered in previous known methods.

▶ Keyword : 유전자 조절망(Gene Regulatory Network), 방향성 비순환 그래프(Directed Acyclic Graph)

---

• 제1저자 : 홍성룡

• 접수일 : 2007.10.29, 심사일 : 2007.11.7, 심사완료일 : 2007.11.20.

\* 한국외국어대학교 컴퓨터및정보통신공학부

※ 이 연구는 2007학년도 한국외국어대학교 교내학술연구비의 지원에 의하여 이루어진 것임

## I. 서론

일반적으로 하나의 유전자는 또 다른 유전자의 단백질과 프로모터 영역에서 Binding함으로써 그 유전자의 발현에 영향을 미칠 수 있다. 이러한 두 유전자간의 상호 조절 작용을 유전자 조절망이라 하며 유전체의 핵심적인 기능을 보다 간결하게 표현하는 조절망을 설계할 수 있다. 따라서 이러한 Microarray chip기술로 유전자의 발현 데이터를 대량으로 얻고 다양한 실험 조건하에 유전자의 발현 양상을 관찰하고 유전자간의 조절 관계를 분석 및 예측 할 수 있게 되었다 [1][2][3][4]. 대표적인 설계 방법으로는 time-series approach [2]와 steady-state approach[1][3][4] 가 있으며 본 논문에서는 steady-state data 즉, knock-out data를 이용한 steady-state approach을 적용하여 유전자 조절망을 재구성하고 기존 방법을 개선하여 보다 정확한 결과를 예측할 수 있는 알고리즘을 제시한다.

본 논문을 구성하는 내용은 다음과 같다. 2절에서는 본 논문과 관련된 생물학적인 배경 지식을, 3절에서는 본 논문의 관련 연구로서 유전자간 조절망을 설계하는 방법으로 선택한 steady-state approach, 유전자간 상호 매트릭스 (Interaction Matrix) 구성 방법, redundant network 설계 방법, S.Cerevisiae Dataset과 실험 방법을, 4절에서는 redundant network에서 중복된 arc를 제거하여 간소화된 조절망인 parsimonious regulatory network를 설계하는 알고리즘으로 strongly connected component를 적용하여 condensation 과정을 수행하여 reduced graph인 DAG(Directed Acyclic Graph) 즉, parsimonious regulatory network를 얻는 내용을, 5절에서는 S.Cerevisiae를 이용한 실험 및 결과 분석으로 알고리즘의 적용 결과를 dot로 도식화하고 마지막으로 6절에서는 결론과 향후 발전 방향을 제시한다.

## II. 관련 연구

### 2.1 배경 지식

본 논문에서는 knockout data를 이용한 steady-state approach를 적용한다[4]. steady-state approach 방법은 임의의 유전자를 제거했을 때 다른 유전자들의 발현에 미치는 영향을 측정하는 방법으로 즉, 다른 유전자들이 발현할 때 임

의 한 gene을 제거시킴으로써 제거하기 전과 후의 다른 유전자들의 발현 level 결과를 보다 명확하게 예측할 수 있는 접근 방법이다. 효모의 일종인 S.Cerevisiae의 데이터에서 임의로 추출한 5개의 유전자로 예를 들면 표 2-1과 같다 [4][5].

표 2-1. log10 (농도)  
Table 2-1. log10 (Intensity)

	ade1	aep2	ald5	anp1	ard1
ade1	0	-0.06	0.07	-0.12	0.13
aep2	-0.22	0	-0.18	-0.42	-0.22
ald5	0.58	0.41	0	-0.19	0.61
anp1	0.2	-0.06	0.07	0	-0.1
ard1	-0.15	-0.12	0.19	-0.13	0

위의 표 2-1에서 ade1, aep2, ald5, anp1, ard1 이 다섯 개의 유전자와 이들에 대해서 1열은 제거되는 유전자를 의미하고 1행은 1열의 각각의 유전자들을 제거함으로써 expression level의 변화를 보이는 유전자들이며 각각의 수치는 그 발현 level에 로그를 취한 값을 의미한다. 또한 여기서 '+'는 다른 유전자의 발현에 positive 영향을 미치는 경우이고 '-'는 negative 영향을 미치게 된다.

interaction matrix(상호 반응 매트릭스)는 위에서 얻은 로그값에 대해 임계값을 적용하여 positive이면 enhancer의 역할을 하기 때문에 '+', negative이면 repressor의 역할이므로 '-'로 구성된 매트릭스이다. 표 2-1에서 임계값을 -0.2~+0.2로 정하였는데 이는 즉 로그값이 -0.2보다 더 작으면 '-'로 정하고 로그값이 +0.2보다 더 크면 '+'로 정한다. 만약 임계값을 적용하지 않았을 경우 모든 유전자간의 매우 작은 영향까지도 모두 고려해야 하기 때문에 조절망이 더 복잡해진다. 따라서 표 2-1에 임계값을 적용한 결과는 아래의 표 2-2와 같다.

표 2-2. 상호반응 매트릭스  
Table 2-2. Interaction Matrix

	ade1	aep2	ald5	anp1	ard1
ade1					
aep2	+			+	+
ald5	-	-			-
anp1	-				
ard1					

여기서 위의 표 2-1과 표 2-2의 동일한 셀을 비교하면 부호의 방향이 반대이다. 이는 각 유전자를 제거했을 경우에 다른 유전자에 positive 영향을 주었다면 실제 제거하지 않았을 경우에는 repressor로서의 영향을 주었기 때문이며 이와 반대로 유전자를 제거했을 경우 다른 유전자에 negative 영향을 주었다면 실제 제거하지 않았을 경우에는 enhancer로서의 영향을 주었기 때문이다. 또한 표 2-1에서의 대각선 셀의 로그값에 대해서는 수치가 존재하지만 임계값과는 상관없이 interaction matrix에 포함하지 않았다. 이는 유전자 각각에 대해 자기 자신에게는 주는 영향을 고려하지 않는다는 의미이다.

redundant network란 interaction matrix로부터 '+', '-'를 고려하여 '+'는 정상적인 화살표로, '-'는 앞이 막힌 화살표로 나타낸 조절망을 말한다. 즉 정상적인 화살표는 positive 영향을 '-'는 negative 영향을 의미하며 화살표의 방향은 '+' 또는 '-'의 영향을 주는 방향을 의미한다.

### III. S.Cerevisiae Dataset과 실험 방법

#### 3.1 S.Cerevisiae Dataset

본 논문에서 사용된 S.Cerevisiae 데이터는 Rosetta Inpharmatics(6)에서 구하였으며 효모의 일종인 S.Cerevisiae의 내부적으로는 global mRNA 발현 하위 집합들을 포함하고 있다. 우선 한 쪽 집합은 Control Set이라 부르는 정상적인 야생 S. Cerevisiae를 63번 샘플링한 knockout mRNA 발현 데이터를 제공하고 다른 집합은 Perturbation Set인 정상적인 효모와 동일한 조건에서 배양한 287개의 유전자에 대해서 각각 knockout된 287개의 S. Cerevisiae의 발현 양을 측정된 수치를 제공하고 있다(7). 하지만 데이터를 살펴본 결과 몇 개의 유전자 명칭과 측정 수치가 일치하지 않는 경우와 로그값의 표시 오류로 인해서 위의 287개의 중 248개만을 실험 데이터로 사용하였다.

##### 3.1.1 실험 방법

먼저 S.Cerevisiae 데이터에서 각각의 유전자 실험에 대한 값들 중에서 log<sub>10</sub>(Intensity) 값만을 따로 추출하여 여기에 임계값을 적용하였다. 임계값은 ±0.1 ~ ±1.0 까지 0.1씩 증가시키면서 로그값을 필터링하여 ±에 따라서 ±1로 정하였다. 앞서 steady-state approach 방법에서의 예는 단지 '±'을 사용하였으나 본 논문에서 제안하는 redundant

network 제거 알고리즘에 이 데이터를 이용하기 위해서 각각의 임계값에 따라서 positive 영향을 주면 '+1'로 negative 영향을 주면 '-1'로 정하였다. 이 데이터를 이용해서 parsimonious regulatory network를 설계하기 위한 준비가 된 셈이다.

실험은 1차 필터링한 데이터에 직접적인 경로를 삭제하기 위해서 redundant network 제거 알고리즘과 transitive closure 알고리즘을 적용하여 interaction matrix를 얻을 수 있는데 여기서 만약 입력 데이터에 사이클이 존재한다면 parsimonious regulatory network를 구성할 수 없기 때문에 이 부분에서 strongly connected component를 적용한 condensation 과정을 수행해야 한다. 이 후 DAG(Directed Acyclic Graph)를 설계한 후 이 데이터를 입력으로 실험을 수행할 것이다.

#### 3.2 Path와 Cycle

##### 3.2.1 Indirected Path를 고려한 개선 사항

조절망을 설계하는데 있어서 그림 3-1과 그림 3-2처럼 제약적인 경우를 살펴보면 다음과 같다.

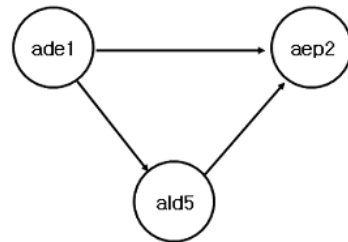


그림 3-1. 제약적인 경우 1  
Figure 3-1. Limitation 1

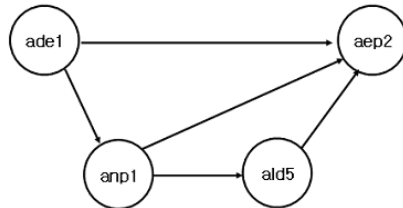


그림 3-2. 제약적인 경우 2  
Figure 3-2. Limitation 2

그림 3-1에서 유전자 ade1은 유전자 aep2에 두 방향으로 positive 역할을 하고 있다. 즉, 직접적으로 enhancer의 영

향을 미치는 정상적인 화살표와 또 하나는 유전자 *ade1*이 유전자 *ald5*를 경유하여 유전자 *aep2*로 간접적인 enhancer의 영향을 미치고 있다. 그림 3-2에서는 유전자 *ade1*이 *anp1*을 거쳐서 *aep2*에 enhancer의 영향을 주고 유전자 *anp1*은 *ald5*를 거쳐서 *aep2*에 enhancer의 영향을 주고 있다. 즉, 기존의 Andreas Wagner의 논문[8]에서는 조절망을 재구성하는 경우 반드시 중간에 경유하는 유전자의 개수가 오로지 하나인 경우 그리고 반드시 positive인 경우에서만 간접적인 영향을 고려하였다. 그러나 그림 3-3과 같이 유전자 *ade1*이 *aep2*에 직접적으로 그리고 또 다른 유전자인 *ald5*를 거쳐서 간접적으로 영향을 주고 있다. 이러한 경우 Wagner의 알고리즘은 올바르게 동작하지 않는다.

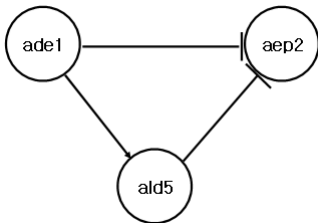


그림 3-3. Wagner 방법의 문제점 1  
Figure 3-3. Problem 1 in Wagner's method

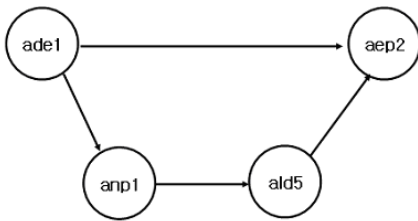


그림 3-4. Wagner 방법의 문제점 2  
Figure 3-4. Problem 2 in Wagner's method

또한 위의 그림 3-4를 보면 유전자간의 주고받는 영향이 모두 positive일지라도 유전자 *ade1*이 *aep2*에 직접적으로 enhancer의 영향을 주며 간접적으로는 유전자 *anp1*를 거쳐 유전자 *ald5*를 경유하여 유전자 *aep2*에 enhancer의 영향을 미치고 있다. 이러한 경우에도 Wagner의 알고리즘은 올바르게 동작하지 않는다.

### 3.2.2 Cycle이 발생하는 경우

그림 3-5를 살펴보면 유전자 *ald5*는 *aep2*, *anp1*, *ade1*을 거쳐서 *ard1*에 enhancer의 영향을 주고 있으며 이 때 세 유전자 *aep2*, *anp1*, *ade1*를 거쳐서 가는 간접적인 경로가 존재하지만 유전자 *ald5*, *aep2*, *anp1*, *ade1* 사이에는 cycle이 발생하게 된다. 그림 3-5에서는 모두 positive인 경우지만 positive와 negative의 영향이 함께 존재한다면 유전자 *ald5*, *aep2*, *anp1*, *ade1*사이의 cycle이 몇 번 반복하는지에 따라 마지막 *ard1*에 미치는 영향이 positive일 수도 negative일 수도 있다. 따라서 4절에서는 조절망을 재구성함에 있어서 그림 3-5와 같이 유전자 *anp1*, *ald5*와 같이 중간에 거쳐서 가는 유전자가 2개 이상인 경우와 그림 3-6처럼 cycle이 있는 경우를 고려하여 Wagner 방법의 문제점을 보완하고 각 유전자간 '±'를 적용하고 transitive closure, strong connected component 그리고 directed acyclic graph 개념을 도입하여 사이클이 없는 parsimonious regulatory network를 설계하고 보다 확실한 결과를 예측할 수 있는 새로운 방법을 제안한다.

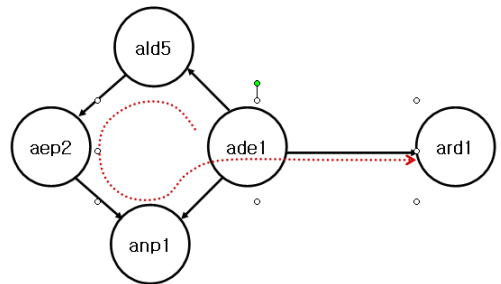


그림 3-5. Cycle이 있는 경우  
Figure 3-5. In case cycles exists

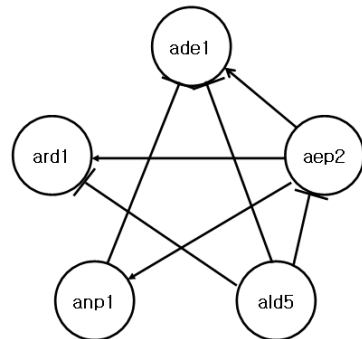


그림 3-6. Redundant Network 예제  
Figure 3-6. Redundant Network Example

## IV. S.Cerevisiae를 이용한 유전자 조절망의 재구성

### 4.1 Parsimonious Regulatory Network

상호 반응 매트릭스에서 각 유전자간의 모든 관계를 도식화하면 매우 복잡해진다. 따라서 redundant network에서 중복되는 관계를 제거하여 parsimonious regulatory network을 구성할 수 있는데 여기에 적용할 수 있는 알고리즘은 다음과 같다.

표 4-1. Parsimonious Regulatory Network 알고리즘  
Table 4-1. Parsimonious Regulatory Network Algorithm

<ol style="list-style-type: none"> <li>For each arc A from U to V on G if arc A is negative and if there exists an alternative path from U to V where number of negative arcs on the path from U to V is odd remove arc A from G</li> <li>For each arc A from U to V on G if arc A is positive and if there exists an alternative path from U to V where number of negative arcs on the path from U to V is even remove arc A from G</li> </ol>
---

U, V : 유전자  
A : arc  
negative, repressor : 유전자의 발현을 억제한다  
positive, enhancer : 유전자의 발현을 촉진한다.

표 4-1에서 U와 V는 유전자를, A는 arc를 의미한다. 유전자의 발현을 억제(repressor)하는 역할을 한다면 negative영향을 주는 것이며 촉진(enhancer)하는 역할을 한다면 positive영향을 주게 된다. 또한 표 4-1의 1에서 유전자 U에서 V로 가는 arc A에 대해 억제를 하며 negative arc의 개수가 홀수인 다른 선택경로(alternative path)가 존재하면 arc A는 삭제 가능하다. 표 4-1의 2에서도 역시 arc A가 positive 영향을 주고 negative arc의 개수가 짝수인 U에서 V로 가는 선택적인 경로(alternative path)가 존재한다면 arc A는 삭제 가능하다.

### 4.2 Redundant arc의 제거

위의 redundant network 제거 알고리즘에서 선택적인

경로(alternative path)를 얻기 위해서는 transitive closure 알고리즘을 적용해야 한다. transitive closure 알고리즘은 그림 4-1과 같이 거쳐서 갈 수 있는 모든 정점을 직접적인 arc로 연결한 그래프를 말한다. 모든 정점에 대해 직접적인 arc를 구하게 되면 거의 모든 정점이 인접하는 정점으로 바뀌게 되고 이는 인접 행렬로 표현하는 것이 유리하며 또한 interaction matrix와 거의 같은 개념으로 설명할 수 있다.

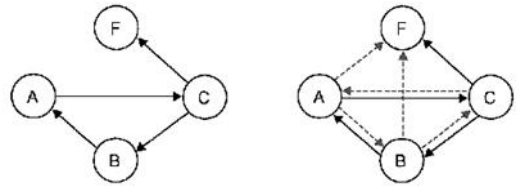


그림 4-1. Redundant arc  
Figure 4-1. Redundant arc

본 논문에서 제안하는 것은 parsimonious regulatory network를 얻을 수 있는 알고리즘이다. 따라서 임의의 유전자에 대한 모든 직접적인 영향을 주는 arc를 얻으면 거의 모든 유전자간의 관계는 인접한 노드의 관계와 동일하다. 여기에 이 두 유전자 사이에 다른 유전자를 경유하는 간접적인 arc를 조사하여 만약 간접적인 화살표가 있다면 직접적인 영향을 주는 arc는 삭제해도 상관없기 때문에 제거한다.

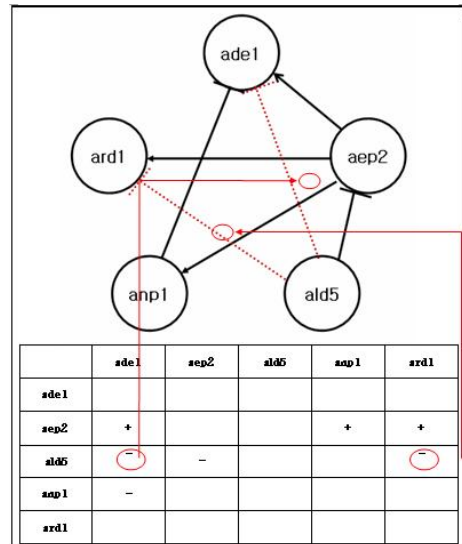


그림 4-2 Interaction Matrix와 Regulatory Network의 관계  
Figure 4-2. Relationship between Interaction Matrix and Regulatory Network

위의 그림 4-2에서 보듯이 두 유전자 ald5와 ade1, ald5와 ard1사이의 negative 직접 경로를 그리고 aep2와 ade1 사이의 positive 직접 경로를 삭제할 수 있는데 유전자 ade1에 대한 ald5의 직접적인 경로의 영향은 '-'이고 간접적인 경로의 영향 역시 '-'이므로 이는 다른 유전자를 경유하는 간접적인 경로가 존재한다는 것이기 때문에 직접적인 경로는 삭제한다. 실제 삭제 작업은 유전자 ade1에 대한 ald5의 직접 경로의 값을 0으로 바꾸는 것이다. 또한 ard1에 대한 ald5의 직접적인 경로 역시 '-'이고 다른 유전자 aep2를 경유하는 간접 경로가 '-'이므로 직접 경로의 값을 0으로 바꿈으로써 역시 삭제 가능하다. 따라서 두 유전자 ald5와 ade1, ald5와 ard1 사이의 negative인 직접 경로를 삭제하면 그림 4-3과 같이 parsimonious regulatory network를 설계할 수 있다.

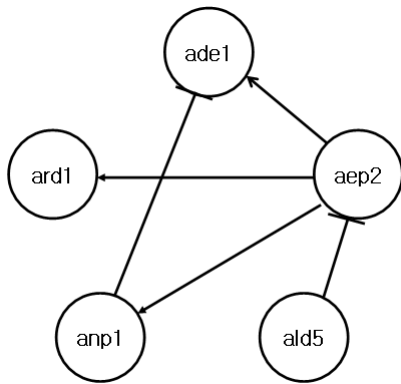


그림 4-3. Parsimonious Regulatory 네트워크  
Figure 4-3. Parsimonious Regulatory Network

### 4.3 Strongly Connected Component

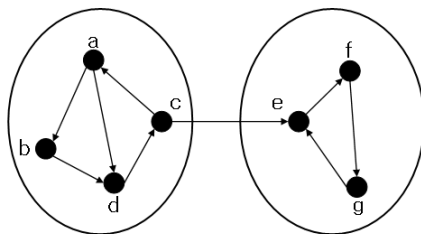


그림 4-4. 강한 연결 요소  
Figure 4-4. Strong Connected Components

방향 그래프  $G = (V, E)$ 에 대해서 임의의 두 vertex을 쌍으로 하는 서로 연결되어 있는 vertex들의 maximal set을

그래프 G의 strongly connected component라고 하며 어떤 digraph가 단 하나의 strongly connected component로 이루어져 있을 때 그 digraph는 strongly connected되었다고 말한다. 그림 4-4에서 {a, b, c, d}와 {e, f, g}가 strongly connected component이다. 여기서 strongly connected component 개념을 알아보는 이유는 입력 데이터 중에서 사이클이 생기는 경우가 있기 때문이다. 만약 유전자 사이에 사이클이 존재하면 이 사이클을 경유하는 횟수에 따라서 negative 역할을 할 수도 있고 positive 역할을 할 수도 있다. 또한 S.Cerevisiae 데이터에는 실험에 관한 여러 가지 로그값들만 나타나있고 반드시 두 유전자 사이에만 사이클이 존재한다는 보장도 없으며 많은 유전자 사이에 어느 정도의 사이클이 존재하는지 알 수 없기 때문에 사이클이 존재하는 유전자 사이에 strongly connected component 개념을 도입하였다. 사이클이 존재하는 데이터에 대한 조절망을 도식화하면 그림 4-5와 같다. 앞의 3절에서 예를 들었던 동일한 유전자이다. 단지 유전자 aep2에서 ald5로 negative 영향을 주는 화살표를 하나 추가 하였다. 이로써 aep2와 ald5 사이로 경유하는 횟수에 따라 다른 유전자에 미치는 화살표의 모양이 달라진다.

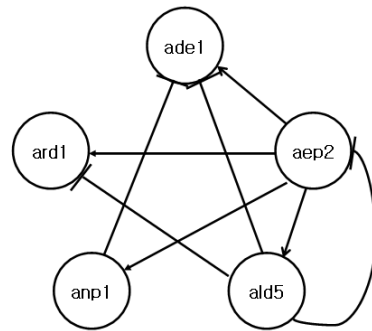


그림 4-5. 싸이클이 존재하는 경우  
Figure 4-5. In case cycles exists

따라서 aep2와 ald5 이 두 유전자를 하나의 component로 구성하기 위해 interaction matrix와 조절망을 다시 작성한다. 앞에서 보았던 interaction matrix와는 조금 달리 구성하는데 이전의 interaction matrix에서의 입력 데이터 값은 "direct positive", "indirect positive", "indirect negative"를 의미하였으나 지금 만드는 interaction matrix의 입력 데이터 값은 "direct positive", "direct negative", "indirect positive", "indirect negative"로 세 개의 부분에서 네 개의 부분으로 늘어난다. 이는 component 개념을 적용하기 때문에 두 유전자를 하나의 유전자로 간주하

게 된다. 따라서 기존의 두 유전자는 하나로 합쳐지고 서로간의 화살표는 삭제되어 이 component에 속하는 각각의 유전자와 다른 유전자 사이의 모든 영향을 고려해야 하기 때문에 직접적인 경로가 두 개가 존재할 수도 있다. 또한 그림 4-6을 살펴보면 component를 구성하는 각각의 유전자와 다른 유전자간의 indirect path가 생길 수 있다.

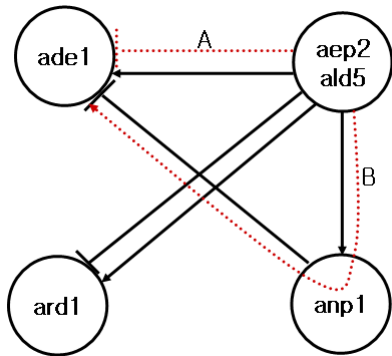


그림 4-6. Component 축약 후의 Indirected Path가 구성된 경우  
Figure 4-6. In case indirected paths exists after condensing components

그림 4-6과 같이 strongly connected component를 구성한 후에 component인 "aep2+ald5"와 다른 유전자들 사이의 모든 path를 고려하면 간접적인 path B가 생성된다. 따라서 component "aep2+ald5"와 ade1사이의 직접적인 path인 A를 삭제해야 한다.

#### 4.4 Directed Acyclic Graph

DAG(Directed Acyclic Graph)는 트리와 같이 사이클을 포함하지 않는 방향 그래프를 의미한다. 앞에서 재구성한 변형된 interaction matrix를 이용하여 재설계한 조절망을 도식화 하면 아래와 같다. 사이클이 존재하는 두 유전자 aep2와 ald5를 컴포넌트로 재설정하여 사이클이 없는 directed acyclic graph를 얻을 수 있다. 이를 바탕으로 기존의 interaction matrix에 직접적인 간선을 하나 더 추가하여 변형된 interaction matrix를 얻을 수 있는데 이 데이터를 입력으로 하여 앞서 보았던 parsimonious regulatory network 알고리즘과 transitive closure 알고리즘을 적용하면 그림 4-7과 같은 결과를 도출할 수 있다. 여기서 변형된 interaction matrix에 대한 두 알고리즘 역시 동시에 두 개의 직접적인 arc가 존재할 수 있기 때문에 입력 데이터에 맞도록 수정이 필요하다.

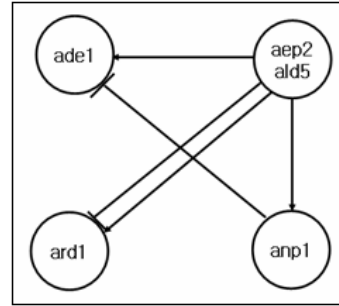


그림 4-7. Parsimonious Regulatory 네트워크  
Figure 4-7. Parsimonious Regulatory Network

### V. S.Cerevisiae 데이터를 이용한 실험 및 결과 분석

본 논문에서는 S.Cerevisiae 효모 데이터에 [6] 대해 로그값(log10.Intensity)을 ±0.1~1.0까지 0.1간격으로 임계치를 적용한 1차적인 필터링 데이터를 얻었다. 이 중 287개의 유전자 실험에 대해서 이름 매칭, 로그값 미지정 문제로 인해 몇몇 개의 유전자를 제외한 248개의 유전자를 대상으로 5개, 10개, 50개, 100개, 150개를 선택하여 Graphviz사의 dot(9)[10]를 적용하였다. 150개 이상의 유전자에 대한 결과는 maximum supported image dimension 이 65500 픽셀이 한계이므로 확인할 수 없었다. 지금까지 예로든 5개의 유전자의 DAG를 도식화하면 그림 5-1과 같다. 그림 5-2는 50개의 유전자를 이용하여 구성한 유전자 조절망이다.

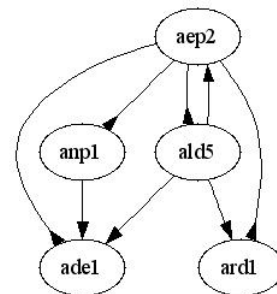


그림 5-1. 유전자 5개를 이용한 S.Cerevisiae 유전자 조절망  
Figure 5-1. S.Cerevisiae gene regulatory network using 5 genes

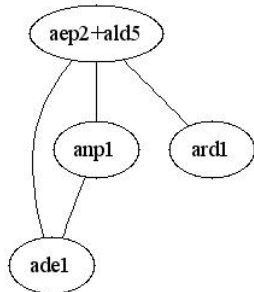


그림 5-2. Strongly Connected Component의 병합  
Figure 5-2. Merging Strongly Connected Components

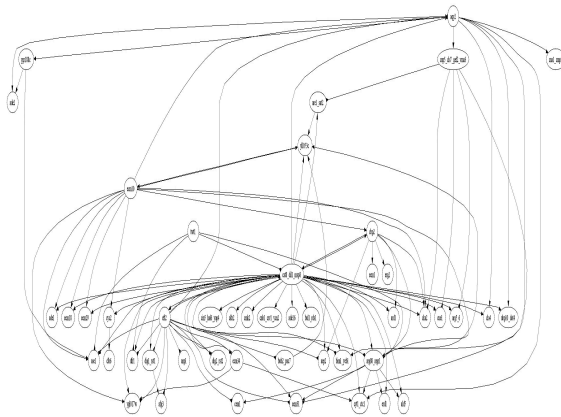


그림 5-3. 유전자 50개를 이용한 S.Cerevisiae 유전자 조절망  
Figure 5-3. S.Cerevisiae gene regulatory network using 50 genes

## VI. 결론

유전자 사이의 negative, positive 영향을 미치는 관계를 도식화한 것을 조절망(Regulatory Network)라 한다. 이 조절망을 설계하는데 있어 대표적으로 time-series approach와 knockout data를 이용한 steady-state approach 방법이 있다. 본 논문에서는 다량의 유전자에서 임의의 유전자를 제거 시킴으로써 그 유전자가 존재할 때와 제거되었을 때 다른 유전자의 발현에 미치는 영향을 보다 확실하게 알 수 있는 steady-state approach 방법으로 접근하였다. 또한 knockout 데이터를 이용한 조절망을 구성하는 방법 중에서 '+'와 '-'를 고려하여 1차적으로 parsimonious regulatory network을 구성하였다. 기존의 방법(8)은 parsimonious network의 구성에서 중요한 indirect path 상의 arc의 향상자(enhancer)와 억제자(repressor)의 역할을 고려하지 않았다. 또한 자동조절(auto-regulation)에 의해서 발생하는 사이

클 문제를 고려하지 않았다. 본 논문에서는 이러한 문제점을 해결하는 방법을 제안하였고 Saccharomyces Cerevisiae의 gene knock-out 데이터에 적용하여 Cerevisiae 유전자 조절망을 구성하였다. 마이크로어레이 실험 데이터로부터 계산학적 방법으로 얻어진 유전자 조절망의 생물학적 정확도 또는 생물학적 의미는 생물학 분야 연구자와의 공동 연구에 의해 검증될 수 있다. 향후 이 분야에 대한 학제간 연구가 필요하다.

## 참고문헌

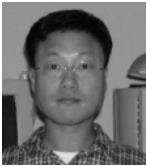
- [1] Y. Gong and Z. Zhang, "Alternative Pathway Approach for Automating Analysis and Validation of Cell Perturbation Networks and Design of Perturbation Experiments," Ann. N.Y. Acad. Sci., 1115(1):pp.267-285, 2007.
- [2] M. Bansal, G. D. Gatta, and D. di Bernardo, "Inference of gene regulatory networks and compound mode of action from time course gene expression profiles," Bioinformatics, 22(7):pp.815-822, 2006.
- [3] J. Tegner, M. K. S. Yeung, J. Hasty, and J. J. Collins, "Reverse engineering gene networks: Integrating genetic perturbations with dynamical modeling," PNAS, 100(10):pp.5944-5949, 2003.
- [4] Steen Knudsen, A Biologist's Guide to ANALYSIS OF DNA MICROARRAY DATA, Wiley & Sons, 2001.
- [5] Andreas D. Baxeavanis, B. F. Francis Ouellette, Bioinformatics : A Practical Guide to the Analysis of Genes and Proteins Second Edition, pp.393-410, Wiley & Sons, 2001.
- [6] <http://www.rii.com> Rosetta Inpharmatics
- [7] Farkas et al. "The topology of the transcription regulatory network in the yeast, Saccharomyces Cerevisiae," Physica A 318, pp.601-612., 2003.
- [8] Wagner, A. "How to reconstruct a large genetic network from n gene perturbations in fewer

than n2 easy steps," Mol. Bio. Evol., 17, pp.1183-1197, 2001.

[9] <http://www.graphviz.org>

[10] <http://search.cpan.org>

## 저 자 소개



손 기락

서울대학교 계산통계학과 학사.

서울대학교 계산통계학과 석사.

미국, Univ. of California, Santa Cruz,

전산학 박사.

전자통신연구원 선임연구원.

현재 한국외국어대학교 컴퓨터및정보통신공학부  
교수



홍 성룡

한국외국어대학교 컴퓨터공학전공 학사

한국외국어대학교 대학원 컴퓨터공학및정보통신  
공학과 졸업, 석사