Research Article

Modeling Gene Networks in *Saccharomyces cerevisiae* Based on Gene Expression Profiles

Yulin Zhang, Kebo Lv, Shudong Wang, Jionglong Su, and Dazhi Meng

1College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao, Shandong 266590, China
2School of Mathematical Sciences, Ocean University of China, Qingdao 266100, China
3College of Computer and Communication Engineering, China University of Petroleum, Qingdao, Shandong 266580, China
4Department of Mathematical Sciences, Xi’an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123, China

Correspondence should be addressed to Kebo Lv; kewave@ouc.edu.cn

Received 14 July 2015; Revised 14 October 2015; Accepted 16 November 2015

Academic Editor: Reinoud Maex

Copyright © 2015 Yulin Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Detailed and innovative analysis of gene regulatory network structures may reveal novel insights to biological mechanisms. Here we study how gene regulatory network in *Saccharomyces cerevisiae* can differ under aerobic and anaerobic conditions. To achieve this, we discretized the gene expression profiles and calculated the self-entropy of down- and upregulation of gene expression as well as joint entropy. Based on these quantities the uncertainty coefficient was calculated for each gene triplet, following which, separate gene logic networks were constructed for the aerobic and anaerobic conditions. Four structural parameters such as average degree, average clustering coefficient, average shortest path, and average betweenness were used to compare the structure of the corresponding aerobic and anaerobic logic networks. Five genes were identified to be putative key components of the two energy metabolisms. Furthermore, community analysis using the Newman fast algorithm revealed two significant communities for the aerobic but only one for the anaerobic network. David Gene Functional Classification suggests that, under aerobic conditions, one such community reflects the cell cycle and cell replication, while the other one is linked to the mitochondrial respiratory chain function.

1. Introduction

The difference between aerobic and anaerobic conditions at the molecular level has drawn considerable attention in last twenty years. Man and Pilpel [1] found that the transcription rate of mitochondrial gene under aerobic conditions is significantly higher than that under anaerobic conditions, while glycolytic genes are more active under anaerobic conditions. Hou et al. [2] found that under anaerobic conditions SPT3 and SPT15 are overexpressed, which may not only enhance resistance to ethanol and stress but also upregulate the fermentation transcription factors in *Saccharomyces cerevisiae*. Jiang et al. [3] found that, under anaerobic condition, mitochondrial function is weakened while fermentation capacity is enhanced in *Saccharomyces cerevisiae*. This suggests competition between aerobic and anaerobic metabolisms as a result of evolution.

A fundamental goal of genomics is to understand how gene regulatory networks actually give rise to cellular phenotypes. Modeling and reconstruction of gene regulatory networks based on high-throughput data have become one of the most common goals of systems biology. Diverse models of gene regulatory networks have been developed such as Boolean network [4–6], probabilistic Boolean network (PBN) [7, 8], and Bayesian network [9–12]. The main objective of these network models is to study the logical interactions of genes based on large-scale microarray data and further get meaningful biological information. Recently using network method to study gene interactions of *Saccharomyces cerevisiae* has attracted the interest of several researchers [13–16]. For example, Zhang et al. [14] examined the integrated gene interaction network of *Saccharomyces cerevisiae* and found many enriched multicolor network motifs corresponding to different biological themes. They concluded that significantly enriched motifs in the network are often signatures of network themes, higher-order network structures that correspond to biological phenomena. Lee et al. [15] modified a probabilistic functional gene network of the baker’s yeast,
Saccharomyces cerevisiae, and experimentally verified the function of the yeast RNA binding protein PuF in 60S ribosomal subunit biogenesis. Hu et al. [16] profiled transcriptional responses in Saccharomyces cerevisiae strains with individual deletions of 263 transcription factors. Then they reconstructed a functional transcriptional regulatory network between these transcription factors and analyzed the enrichment of promoter motifs on these transcription factors.

In 2005, Bowers et al. [17] introduced a computational approach called logic analysis of phylogenetic profiles (LAPP), which identified detailed logic relationships among gene triplets on the basis of genomic data. This method may be used for functional annotation of proteins and genes and for designing biochemical experiments to elucidate biological mechanisms. Lately, further progress has been achieved on the theory and application of higher logic. Zhang et al. [18] described a three-way gene interaction model that captures the dynamics of coexpression relationships between two genes. Shoemaker and Panchenko [19] proposed ways to address the defects of the LAPP method, such as high computational complexity, strong dependence on information spectrum, and the uncertainty of homology detection at large genetic distances. Sprinzak et al. [20] detected coordinated regulation of multiple protein complexes using logic analysis of gene expression data and identified protein complexes by mapping specific kinds of gene triples to multicomplexes triplets. Notably, the LAPP method is related to stochastic logic. Modeling approaches using stochastic logic, such as stochastic Boolean network (SBN) and stochastic multiplicity-valued network (SMN), have been already proposed [21–24].

In this paper, we focus on the construction, analysis, and comparisons of structural characteristics of logic networks inferred from the gene expression profiles of Saccharomyces cerevisiae under aerobic and anaerobic conditions. Firstly, gene expression profiles are discretized into multiple values. Secondly, the logical AND, OR, and NOT operators are given by algebraic formulation on multiple values. Down- and upregulation self-entropy as well as joint entropy are used to compute the uncertainty coefficient. Four parameters of the logic network are generalized to more complex networks for our purpose. Putative regulator genes of respiratory mechanisms are identified by contrasting the differences of the four structural parameters. Lastly, the Newman fast algorithm is used to discover the community structures of the logic network. We find that the gene logic network under aerobic condition (aerobic network) has two significant communities while the anaerobic network only has one in this framework. Furthermore, David Gene Functional Classification (http://david.abcc.ncifcrf.gov) [25] reveals the possible biological function of these communities.

2. Materials and Methods

2.1. Expression Data. The expression profiles of Saccharomyces cerevisiae strains under aerobic and anaerobic conditions (Table 1). Such a large number of genes incur an intractable computational complexity, so we select candidate genes based on the Wilcoxon rank sum test [26] as follows. Let $X_j = (x_{j1}, x_{j2}, \ldots, x_{jn})$, $Y_j = (y_{j1}, y_{j2}, \ldots, y_{jn})$ denote $n_0$ and $n_1$-sized random sample of expression of gene $j$ under aerobic and anaerobic conditions, respectively. The Wilcoxon rank sum test statistic is derived from the concatenation of $X_j$ and $Y_j$, which results in vector $Z_j$. This is done by first sorting all points in $Z_j$ by ascending order, obtaining the sets of ranks for the $X_j$ and $Y_j$ points, respectively, then calculating the corresponding sums $T_{j0}$ and $T_{j1}$, and finally defining the test statistic as $T_{j1}/n_1 - T_{j0}/n_1$. Under the null hypothesis the expected value of this statistic is zero; otherwise the alternative hypothesis holds, in which case $H_0: T_{j0}/n_0 - T_{j1}/n_1 = 0; H_1: T_{j0}/n_0 - T_{j1}/n_1 \neq 0$, then gene $j$ is called a candidate gene. Setting the significance level to be $10^{-5}$, we obtain 73 candidate genes finally.

Table 1: The data source.

<table>
<thead>
<tr>
<th>Group</th>
<th>Platform</th>
<th>Database</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerobic</td>
<td>GPL90</td>
<td>GSE11452</td>
<td>42</td>
</tr>
<tr>
<td>Anaerobic</td>
<td>GPL90</td>
<td>GSE11452</td>
<td>52</td>
</tr>
</tbody>
</table>

2.2. Methods

2.2.1. Construction of the Gene Logic Network. The expression profile of candidate gene $m$ containing $n$ samples is denoted in this work by the vector $(d_{m1}, \ldots, d_{mp}, \ldots, d_{mn})$. To discretize this vector, we define $x_i = (i - 1)/2k$, $x_i = 1 - (i - 1)/2k$, $i = 1, 2, \ldots, k, k \in \mathbb{N}$, where the tuning parameter $k$ controls granularity on the interval $[0, 1]$. Then for $(d_{m1}, \ldots, d_{mp}, \ldots, d_{mn})$, we set $d_{mp} = x_i$ if $d_{mp} \in [(i - 1)/2k, i/2k)$, $(1 \leq p \leq n)$, and similarly $d_{mp} = x_i$ if $d_{mp} \in [1 - i/2k, 1 - (i - 1)/2k)$, obtaining the discretized expression vector $D = (d_{m1}, \ldots, d_{mp}, \ldots, d_{mn})$. In order to calculate first-order and second-order logical relations using LAPP method, we define the following quantities for this discretized vector:

(1) Downregulation self-entropy is $H_-(D) = -\sum_{i=1}^{n} p_D(x_i) \log(p_D(x_i))$.

(2) Upregulation self-entropy is $H_+(D) = -\sum_{i=1}^{n} p_D(x_i) \log(p_D(x_i))$.

Then self-entropy is $H(D) = H_-(D) + H_+(D)$, where $p_D(x_i)$ and $p_D(x_i)$ are the corresponding frequency of components $x_i$ and $x_i$, respectively, in vector $D$. Considering two vectors $D = (d_{m1}, \ldots, d_{mp}, \ldots, d_{mn})$ and $B = (b_{m1}, b_{m2}, \ldots, b_{mn})$, we define the following joint entropies:

(3) $H_+(D, B) = -\sum_{i=1}^{n} p_{D,B}(x_i, x_i^0) \log(p_{D,B}(x_i, x_i^0))$.

(4) $H_-(D, B) = -\sum_{i=1}^{n} p_{D,B}(x_i, x_i^0) \log(p_{D,B}(x_i, x_i^0))$.

(5) $H_-(D, B) = -\sum_{i,j=1}^{n} p_{D,B}(x_i, x_i^0) \log(p_{D,B}(x_i, x_i^0))$.

(6) $H_+(D, B) = -\sum_{i,j=1}^{n} p_{D,B}(x_i, x_i^0) \log(p_{D,B}(x_i, x_i^0))$. 

20,000 genes in yeast under aerobic and anaerobic conditions for our purpose. Putative regulator genes of respiratory mechanisms are identified by contrasting the differences of the four structural parameters. Lastly, the Newman fast algorithm is used to discover the community structures of the logic network. We find that the gene logic network under aerobic condition (aerobic network) has two significant communities while the anaerobic network only has one in this framework. Furthermore, David Gene Functional Classification (http://david.abcc.ncifcrf.gov) [25] reveals the possible biological function of these communities.
Table 2: Illustration of 2-order logical types.

<table>
<thead>
<tr>
<th>Type</th>
<th>Illustration</th>
<th>The proper function</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C is present iff A and B are both present</td>
<td>( C = A \land B )</td>
<td>( \min (a_i, b_i) )</td>
</tr>
<tr>
<td>2</td>
<td>C is present iff A is absent or B is absent</td>
<td>( C = \neg (A \land B) )</td>
<td>( \max (\neg a_i, \neg b_i) )</td>
</tr>
<tr>
<td>3</td>
<td>C is present iff A is present or B is present</td>
<td>( C = (A \lor B) )</td>
<td>( \max (a_i, b_i) )</td>
</tr>
<tr>
<td>4</td>
<td>C is present iff A is absent and B is absent</td>
<td>( C = \neg (A \lor B) )</td>
<td>( \min (\neg a_i, \neg b_i) )</td>
</tr>
<tr>
<td>5</td>
<td>C is present iff A(B) is absent and B(A) is present</td>
<td>( C = \neg (A \land B), C = (A \land \neg B) )</td>
<td>( \min (\neg a_i, b_i), \min (a_i, \neg b_i) )</td>
</tr>
<tr>
<td>6</td>
<td>C is present iff A(B) is absent or B(A) is present</td>
<td>( C = \neg (A \lor B), C = (A \lor \neg B) )</td>
<td>( \max (\neg a_i, b_i), \max (a_i, \neg b_i) )</td>
</tr>
<tr>
<td>7</td>
<td>C is present iff A and B are both present or A and B are both absent</td>
<td>( C = (A \leftrightarrow B) )</td>
<td>( \max {\max (\neg a_i, b_i), \max (a_i, \neg b_i)} )</td>
</tr>
<tr>
<td>8</td>
<td>C is present iff one of either A or B is present</td>
<td>( C = \neg (A \leftrightarrow B) )</td>
<td>( \max {\min (\neg a_i, b_i), \min (a_i, \neg b_i)} )</td>
</tr>
</tbody>
</table>

And total joint entropy is \( H(D, B) = H_1(D, B) + H_2(D, B) + H_3(D, B) + H_4(D, B) \), where \( p_{D,B}(x_i, x_j) \) are the corresponding frequencies of component \( x_i x_j \) in vectors \( D \) and \( B \), respectively. For discretized vectors \( A \) and \( B \), the uncertainty coefficient \( U \) is defined as

\[
U(B | A) = H(B) + H(A) - \frac{H(A, B)}{H(B)}. 
\]

(1)

This quantity informs on the probability that \( A \) regulates \( B \). Note that for simplicity \( A \) and \( B \) denote not only expression vectors but also the corresponding genes. The first-order logical relationship between genes \( A \) and \( B \) is determined as

\[
U(B | f_1(A)) = H(B) + H(f_1(A)) - \frac{H(B, f_1(A))}{H(B)}, 
\]

(2)

where \( f_1 \) is the proper functions of first-order logic of \( A \) to \( B \). The uncertainty coefficient of \( A, B \rightarrow C \) is

\[
U(C | f_2(A, B)) = H(C) + H(f_2(A, B)) - \frac{H(C, f_2(A, B))}{H(C)}, 
\]

(3)

where \( f_2 \) is one of the proper functions of second-order logic of \( A \) to \( B \). Table 2 lists ten types of proper functions and the corresponding algebraic operations. Using these operations on \( A \) and \( B \), we get \( C \). Table 2 also gives the algebraic representation of three basic operators for multiple values: logical NOT is represented as \( \neg (a_i = x_i^k \text{ or } x_i^l) = (a_i = x_i^k \text{ or } x_i^l) \), logical AND and OR can be represented as \( \min (a_i, b_i), \max (a_i, b_i) \), where \( a_i, b_i \) are the \( i \)th component of \( A, B \).

We normalize the \( U \) value for each gene triplet and database by replacing it with \( U/\max U \), where \( \max U \) is the maximum value in the own database. For simplicity, the normalized values are also denoted by \( U \). The condition

\[
U(C | f_2(A, B)) \preceq \max(U(C | A) + t, U(C | B) + t) 
\]

is used to filter out all gene triplets. The combination requires gene \( C \) to be better predicted from genes \( A \) and \( B \) together than just gene \( A \) alone or gene \( B \) alone. Figure 1 shows an example for a gene triplet, for which the logical AND operation on gene \( A \) and gene \( B \) is denoted by \( A \lor B \rightarrow C \). As in the LAPP method, all such gene triplets, with the corresponding the \( U \) values, give rise to the gene logic networks further studied in our present work.

2.2.2. Structural Parameters Definition of Logic Network. In fact, the logic network can be seen as a directed and weighted network without multiple edges and self-loops. Let \( G = (V, E, W) \) be a logic network with gene node-set \( V \), directed edge-set \( E \), and function \( W: E \rightarrow R \) that assigns each edge \( e \in E \) and weight \( W(e) \in R \). In fact, the edge weights can be interpreted as the uncertainty coefficients that express interaction strength between gene triplets. Various structural parameters have used to study network structure in complex network [27]. For the logic network, structural parameters are detected again including average degree, average path length, average clustering coefficient, and average betweenness to
capture structural characteristics of the logic network from different angles.

1) Average Degree (D). The degree of a node is the number of nodes adjacent to it. The average degree is the average value of the degrees of all nodes. For the logic network, the definitions of in-degree, out-degree need to be reconsidered based on the principle that the sum of in-degree and that of out-degree of all the nodes in a network are equivalent. Based on the principle, in-degree and out-degree can be defined according to second-order logical relationship; if there are k activations of C, then in-degree of C increases by k/2. However, the out-degrees of A and B are determined by the proportion of their contributions to the second-order logical relationship. Here we can assume that the proportion contribution to the second-order logical relationship from A and that from B are always equivalent. In other words, the out-degree increment of A is the same as that of B.

2) Average Clustering Coefficient (C). For a certain node with second-order logical relationships, we need to define doublets of second-order logical relationships to measure the clustering coefficient. A doublet of second-order logical relationships is a combination of two second-order logical relationships with at least one common node.

If the common node is v, we call this second-order logic doublet centered on v. Figure 2 shows all possible second-order logic doubles centered on v. These three types of second-order logic doubles are named according to the different positions of v as “both-in,” “both-out,” and “in-out” doublets. If there are two common nodes in a second-order logic doublet, we call the doublet strong connected. Figure 3 shows all possible strong connected second-order logic doublets centered on v. The number of second-order logic doublets (including both-in, both-out, and in-out doublets) centered on v is denoted by $\sigma_{\text{doub}}(v)$, and the number of strong connected second-order logic doublets is denoted by $\sigma_{\text{sc-doub}}(v)$. The clustering coefficient of node v, denoted by $C_{\text{doub}}(v)$, in a logic network with second-order logical relationships is defined as $C_{\text{doub}}(v) = \sigma_{\text{sc-doub}}(v)/\sigma_{\text{doub}}(v)$. So the average clustering coefficient of network is defined as $C_G = (1/n) \sum_{v \in G} C_{\text{doub}}(v)$.

3) Average Path Length (L). Path and its length should be reconsidered according to the different second-order logic types: AND, OR, and XOR. Take a second-order logical relationship $(A, B) \rightarrow C$, and an arbitrary node in V other than C (say X), the shortest directed path $p_{XC}$, and the distance $d_{XC}$ from X to C are defined as follows.

Case 1 (second-order logic type is AND). Namely, nodes A and B regulate node C cooperatively.

1) X is A or B (say A without loss of generality). The shortest paths from X to B are denoted by $p_{XB}$, if there is at least one directed path from X to B. The directed path $p_{XC}$ from X to C arrives at B through $p_{XB}$ and then the second-order logical relationship to C. The distance $d_{XC}$ from X to C is the total sum of $d_{XB}$ and $d_{(AB)\rightarrow C}$, where $d_{(AB)\rightarrow C}$ is the length of the second-order logical relationship. In our study, $d_{(AB)\rightarrow C}$ is estimated by the reciprocal of the uncertainty coefficient of this second-order logical relationship. On the other hand, if v and B are not connected by a directed path, there is no path from v to C.

2) X is neither A nor B. X is reachable from at most one of A and B. Then there is no directed path starting from X and ending at C. X is reachable from both A and B. Then there is at least one directed path connecting X and C. The distance $d_{XC}$ from X to C is $d_{XC} = \max(d_{XA}, d_{XB}) + d_{(AB)\rightarrow C}$. The directed path $p_{XC}$ first reaches the nearer one of A and B followed by the other and finally C.

Case 2 (second-order logic type is OR). Either A or B can regulate C independently. The distance $d_{AC}$ from A to C is estimated by the probability that activation of C results from A. The next step is to distribute the uncertainty coefficient $U(C \mid f_2(A, B))$ of the second-order logical relationship. Calculate the uncertainty coefficients caused by A and B, denoted by $U(C \mid A)$ and $U(C \mid B)$, respectively, as follows:

$$U(C \mid A) = \frac{1}{2} \cdot U(C \mid f_2(A, B)),$$

$$U(CB) = \frac{1}{2} \cdot U(C \mid f_2(A, B)). (4)$$

On the basis of above obtained uncertainty coefficients, the shortest directed path $p_{XC}$ and the distance $d_{XC}$ from X to C can be determined as follows.

1) X is A or B (say A without loss of generality). There is at least one path starting with X and ending at B.

![Figure 2: (a) Double-in second-order logic doublets centered on i. (b) Double-out second-order logic doublets centered on i. (c) In-out second-order logic doublets.](image)
Choose one of the shortest directed paths from \( X \) to \( B \) randomly, denoted by \( p_{XB} \). The total sum of \( d_{XB} \) and the reciprocal of \( U(C \mid B) \) are denoted by \( d^{(1)}_{XC} \) and \( d^{(2)}_{XC} \), respectively. If \( d^{(1)}_{XC} \) is less than \( d^{(2)}_{XC} \), the shortest directed path \( p_{XC} \) goes directly from \( X \) to \( C \) through the second-order logical relationship, and the distance \( d_{XC} \) equals \( d^{(1)}_{XC} \). Otherwise, \( p_{XC} \) first arrives at \( B \) and then \( C \), and \( d_{XC} \) is equal to \( d^{(2)}_{XC} \). There is no path from \( A \) to \( B \). The path from \( X \) to \( C \) is only the second-order logical relationship, and the distance \( d_{XC} \) equals the reciprocal of \( U(C \mid A) \).

\[ (2) \quad X \text{ is neither } A \text{ nor } B. \text{ There is at least one path starting from } v \text{ towards } A \text{ or } B. \text{ Therefore, the distance from } X \text{ to } C \text{ is } d_{XC} = \min[d_{XA} + d_{(A,B)\rightarrow C}, \min d_{XB} + d_{(A,B)\rightarrow C}]. \text{ The shortest directed path from } X \text{ to } C \text{ is the corresponding path to the choice of distance. Note that when } X \text{ is unreachable from } A \text{ (or } B), d_{XA} \text{ (or } d_{XB}) \text{ is infinite. Neither } A \text{ nor } B \text{ is reachable from } X \text{. The distance from } X \text{ to } C \text{ is infinite and no path connects them in this case.} \]

Case 3 (second-order logic type is XOR). Both \( A \) and \( B \) can activate \( C \) cooperatively or independently. Therefore, XOR type of second-order logical relationship is a combination of AND type and OR type. However, when only one of \( A \) and \( B \) is reachable from \( X \), the condition is the same as OR logic. When both \( A \) and \( B \) are reachable from \( X \), or \( X \) is \( A \) or \( B \), it is the same as AND logic.

According to the definition above, all shortest directed paths and all distances can be found in a logic network. And the average path length of a logic network is defined as \( L_G = \frac{1}{|D|} \sum_{(s,t) \in E} d_{st} \), where \( D \) is the set of ordered pairs of nodes and the distance from the first one to the second one is finite; that is, \( D = \{(s,t) \mid s,t \in V, d_{st} < +\infty\} \) \(|D|\) is the number of the elements in \( D \).

\[ (4) \quad \text{Average Betweenness (B). Betweenness centrality [28] is one indicator of a node's centrality in complex network. It is equal to the number of shortest paths from all vertices to all others that pass through that node. A node with high betweenness centrality has a large influence on the transfer of items through the network, under the assumption that item transfer follows the shortest paths. Let } L_G \text{ be the set of all the shortest paths (allowing more than one shortest paths between two nodes) in the logic network } G. \text{ If node } v \text{ appears at least once in a directed path } p_v \text{ starting with } s \text{ and ending at } t, \text{ then } v \text{ is referred to as intermediate node in this path, denoted by } v \in p_v. \text{ The standard betweenness of } v, \text{ denoted by } C_{\text{Logic}}(v), \text{ is defined as} \]

\[ C_{\text{Logic}}(v) = \frac{\sum_{s \in V \setminus \{v\}} \sum_{t \in V \setminus \{v\}} [p_{st} \mid v \in p_{st}]}{|L_G|}. \]

\[ (5) \]

The standard betweenness centrality ranges from zero to one. Higher betweenness means larger possibility to appear in the shortest paths. It is more important in the structure and information transfer of a network. Therefore betweenness can help to discover crucial nodes that may have significant impacts on the structural characteristics of the logic network.

2.2.3. Community Structures in the Logic Network. Another important feature of complex network is its community structures, which depict the organization of vertices into clusters, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters. The community is a good tool to describe network structures and provides better understanding of network functions. Several researchers [29–31] proposed algorithms to detect community structures. Newman fast algorithm [32] is a greedy modularity algorithm starting from a set of isolated nodes. The links of the original graph are iteratively added such that the largest possible increase of the modularity at each step is achieved. The fast algorithm is used to find the community structure in the logic network. In addition, the inside-to-outside-of-community ratio \( \alpha = n_1/n_2 \) is used to evaluate how close of community connections in the network, where \( n_1 \) is number of edges within a community and \( n_2 \) is the number of edges cooperating the internal and external node in a community.
3. Results

In order to highlight the characteristics of the logic network structures, we contrast the change curves (Figure 4) of four structural parameters along with threshold $t$ from 0.1 to 0.9 between the aerobic network and the anaerobic network with step length 0.1. We find that the average degree, average clustering coefficient, and average betweenness of the aerobic network are greater than those of the anaerobic network for all thresholds. It can be seen that the average path length of the aerobic network is greater than that of the anaerobic network in some threshold ranging from 0.3 to 0.7. The significant changes of parameters mean that the energy metabolism conditions for *Saccharomyces cerevisiae* in aerobic and anaerobic respirations actually differ on the molecular level.

Each node of the logic network corresponds to a different structural parameter. From this, we obtain the degree, clustering coefficient, betweenness, and path length of each node. By calculating and ranking the difference values of the four parameters for each node, we capture the top five gene nodes. In Table 3, $D$-Difference, $L$-Difference, $C$-Difference, and $B$-Difference denote the difference values of the four structural parameters, respectively. Finally we get the intersection of these gene nodes which includes genes ATP6, YIG1, RGI2, BAG7, and COX1. The structural parameters of these genes change significantly comparing the aerobic network with the anaerobic network. For example, some genes are allocated a higher degree in the aerobic network than in the anaerobic network. That is to say, these genes have great contribution to degree structure changes for the two networks, so we define them as the structural key genes.

If gene $A$ connects other genes by a certain second-order logic relationship, then these genes together from one set are denoted by $S_A$. For example, if $C \cap D \rightarrow A$, then $C \in S_A$, $D \in S_A$. If $A \cap F \rightarrow G$, then $G \in S_A$. If the gene of $S_A$ has a particular function, then we predict that gene $A$ also has the same function. David Gene Functional Classification analysis is carried out on $S_A$. The second column of Table 4 shows that the functional annotation of these genes has been detected in David database while the final column lists the predicted functions of them.

Table 5 shows nonisolated nodes, the numbers, and modularity of community structures. Utilizing Newman fast algorithm, we find that the aerobic network has two obvious community structures. The modularity is 0.3756 including 47 and 15 nonisolated nodes. The corresponding ratio of inside and outside of community is 87.3 and 19.8, respectively. There are three community structures in anaerobic network, including 13, 16, and 5 nodes. The corresponding ratio of...
Table 3: Difference of parameters for some genes between the aerobic and the anaerobic network.

<table>
<thead>
<tr>
<th>Gene</th>
<th>COX1</th>
<th>ATP6</th>
<th>COB</th>
<th>BAG7</th>
<th>YIG1</th>
<th>RGI2</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-Difference</td>
<td>22</td>
<td>22</td>
<td>21</td>
<td>20</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Gene</td>
<td>COX1</td>
<td>YIG1</td>
<td>BAG7</td>
<td>BI3</td>
<td>ATP6</td>
<td>RGI2</td>
</tr>
<tr>
<td>L-Difference</td>
<td>12</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Gene</td>
<td>COB</td>
<td>ATP6</td>
<td>COX1</td>
<td>BAG7</td>
<td>YIG1</td>
<td>RGI2</td>
</tr>
<tr>
<td>C-Difference</td>
<td>2.320</td>
<td>2.194</td>
<td>2.183</td>
<td>2.110</td>
<td>2.065</td>
<td>1.910</td>
</tr>
<tr>
<td>Gene</td>
<td>BAG7</td>
<td>YIG1</td>
<td>RGI2</td>
<td>COX3</td>
<td>COX1</td>
<td>ATP6</td>
</tr>
<tr>
<td>B-Difference</td>
<td>0.286</td>
<td>0.263</td>
<td>0.255</td>
<td>0.207</td>
<td>0.186</td>
<td>0.172</td>
</tr>
</tbody>
</table>

Table 4: Biological function of structural key genes.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Biological function annotation</th>
<th>Predicted function</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP6</td>
<td>Mitochondrial membrane ATP synthetase, participating in the respiratory chain</td>
<td>Involved in yeast mitochondrial respiratory chain</td>
</tr>
<tr>
<td>RGI2</td>
<td>Associated with energy metabolism under condition of aerobic respiration</td>
<td>Involved in membrane transport</td>
</tr>
<tr>
<td>COX1</td>
<td>Cytochrome C-oxidase, catalyzing the reduction of oxygen to water in the respiratory chain</td>
<td>Involved in yeast mitochondrial respiratory chain</td>
</tr>
<tr>
<td>BAG7</td>
<td>Signal conduction function, activation of RH01 which can regulate Gsc2p and Fks1p</td>
<td>Involved in yeast mitochondrial respiratory chain</td>
</tr>
<tr>
<td>YIG1</td>
<td>Compiling protein lipid interactions under anaerobic conditions and associated with the production of glycerol metabolism</td>
<td>Involved in yeast mitochondrial respiratory chain</td>
</tr>
</tbody>
</table>

Table 5: Modularity of the logic networks.

<table>
<thead>
<tr>
<th>Group</th>
<th>Nonisolated nodes</th>
<th>Number</th>
<th>Modularity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerobic network</td>
<td>67</td>
<td>2</td>
<td>0.3756</td>
</tr>
<tr>
<td>Anaerobic network</td>
<td>44</td>
<td>3</td>
<td>0.2842</td>
</tr>
</tbody>
</table>

inside and outside of community is 1.8, 6.0 and 0.57 (see Table 6).

4. Conclusions

By David Gene Functional Classification, we predict that the genes ATP6 COX1 BAG7 and YIG1 are involved in the yeast mitochondrial respiratory chain while RGI2 is involved in membrane transportation (Table 4). In fact, ATP6 [33] belongs to the family of genes, which is called the mitochondrial respiratory chain complex. It participates in the respiratory chain and provides information for synthesis of a protein encoded by mitochondrial DNA that is essential for normal mitochondrial function. YIG1 [34] encodes protein and is involved in regulating anaerobic glycerol metabolism in Saccharomyces cerevisiae. Deletion or overexpression of YIG1 significantly affects growth yield or glycerol yield in anaerobic batch cultures. This is consistent with the previously proposed low flux control exerted at the Gpp level. BAG7 [35] is involved in the energy metabolism under aerobic conditions; its expression is induced under carbon limitation and suppressed under high glucose; COX1 [36] belongs to the family of cytochrome C-oxidase, catalyzing the reduction of oxygen to water in the respiratory chain. RGI2 [37] is a protein of unknown function associated with metabolism under conditions of aerobic respiration; its expression is induced under carbon limitation and suppressed under high glucose. It is involved in the control of metabolism and significantly contributes to cell fitness, especially under respiratory growth conditions.

Some biological functions of organism are realized by gene interactions of proteins. These genes are closely linked, thus revealing the community phenomenon of the logic network. We produce the spider diagrams in Figures 5 and 6 using the Pajek software (http://mrvar.fdv.uni-lj.si/pajek/). The aerobic network is shown to have two distinct communities while the anaerobic network has an obvious community. These communities may correspond to certain biological functions. So we apply the David Gene Functional Classification analysis to these communities. The result reveals that, in the first community of the aerobic network, seven genes including 4533 at, 5527 at, 8444 at, 8426 at, 10446 s at, 6116 at, and 6908 at are related to the cell cycle function. Ten genes including 2565 s at, 2498 s at, 3215 f at, 8822 at, 6044 at, 2361 s at, 6590 at, 10644 at, 8431 at, and 2425 at are associated with the DNA transcription. In the second community of the aerobic network, 12 genes including 3975 at, 3970 s at, 3974 at, 3959 at, 4008 at, 3988 at, 6166 at, and 6908 at are related to the cell cycle function. Ten genes including 2565 s at, 2498 s at, 3215 f at, 8822 at, 6044 at, 2361 s at, 6590 at, 10644 at, 8431 at, and 2425 at are associated with the DNA transcription. In the second community of the aerobic network, 12 genes including 3975 at, 3970 s at, 3974 at, 3959 at, 4008 at, 3988 at, 8431 at, and 2425 at are associated with the DNA transcription. In the second community of the aerobic network, 12 genes including 3975 at, 3970 s at, 3974 at, 3959 at, 4008 at, 3988 at, 8431 at, and 2425 at are associated with the DNA transcription.

5. Discussions

For the nonbinary gene expression profiles, this work gives a novel approach in calculating the self-entropy (including downward self-entropy, upregulation self-entropy) and joint entropy for gene vectors. LAPP method is utilized to find
Table 6: Communities of the logic networks.

<table>
<thead>
<tr>
<th>Group</th>
<th>Community</th>
<th>Nodes in community</th>
<th>Ratio of inside and outside of community</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerobic network</td>
<td>Community 1</td>
<td>50</td>
<td>89.3</td>
</tr>
<tr>
<td></td>
<td>Community 2</td>
<td>17</td>
<td>18.8</td>
</tr>
<tr>
<td>Anaerobic network</td>
<td>Community 1</td>
<td>11</td>
<td>1.80</td>
</tr>
<tr>
<td></td>
<td>Community 2</td>
<td>28</td>
<td>6.43</td>
</tr>
<tr>
<td></td>
<td>Community 3</td>
<td>5</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Figure 5: Spider diagram of the aerobic network.

All gene triplets. Furthermore, the gene logic networks are constructed. The relationships between the structure and the function of the network could help us to understand metabolisms of *Saccharomyces cerevisiae* on the molecular level. To analyze the structural difference between two networks, parameters such as average degree, average clustering coefficient, average path length, and average betweenness, which reveal that the second-order logical types have significant differences among the different experimental gene sets, have been generalized to the networks. The differences may provide us with a new idea and some reference to the biologists on their research work. However, how the other analytical methods, such as community structures, can be generalized into the logic networks is still an interesting issue. By applying the Newman fast algorithm to the logic networks, we find that the aerobic network has two significant...
communities, while the anaerobic network has only one significant community. Furthermore the David Gene Functional Classification shows that one of two communities possibly relates to cell cycle and cell replication functional group; the other possibly correlates mitochondrial respiratory chain functional group in the aerobic network.

Conflict of Interests
The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments
The research is supported by Grants 61170183, 10971122, and 61272018 from National Natural Science Foundation of China, Scientific Research Foundation of Shandong University of Science and Technology for Recruited Talents (2014RCJ033). The authors would like to thank Dr. Attila Gulyás-Kovács at the Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, for their help in revising the paper.

References

Figure 6: Spider diagram of the anaerobic network.


