Validation of Gene Regulatory Networks from Protein-Protein Interaction Data: Application to Cell-Cycle Regulation

Iti Chaturvedi[1,2], Meena Kishore Sakharkar[2], and Jagath C. Rajapakse[1]

[1] Bioinformatics Research Center, Nanyang Technological University, Singapore
[2] Adams Lab, MAE, Nanyang Technological University, Singapore
asjagath@ntu.edu.sg

Abstract. We develop a technique to validate large-scale gene regulatory networks (GRN) by comparing with corresponding protein-protein interaction (PPI) networks. The GRN are obtained with Bayesian networks while PPI networks are obtained from database of known PPI interactions. We look for exact matches and then reduced networks by skipping one or more genes in GRN. We demonstrate our technique on expression profiles of differentially expressed genes in the *S. cerevisiae* cell cycle. We validate GRNs against a merged database of 53235 genes. The precisions of GRN obtained over all genes were from 0.82 to 0.95 in all the phases. In particular we realized that one-skip and two-skip model significantly improved accuracy of the GRN of different phases of cell cycle.

Keywords: Dynamic Bayesian networks, gene regulatory networks, genetic algorithms, protein-protein interactions.

1 Introduction

A *protein-protein interaction network* (PPIN) has protein as nodes and the edges can be signaling, regulatory and biochemical interactions of the proteome. However, a *Gene Regulatory Network* (GRN) shows interaction of DNA segments of the genome with other substances of the cell, which results in regulating rates at which genes are transcribed to mRNA. This high throughput data has a large scope for organization in context of disease and biological function [1]. There is a need to explain the cellular machinery of a GRN in a systems biology perspective as seen by a PPIN. A common representation of GRN is a '*pathway model*' , a graph where vertices represent genes (or larger chromosomal regions) and arcs represent casual pathways. A vertex can either be off/normal or on/abnormal. Bayesian networks (BN) have recently become popular in deriving and deciphering GRN [2] and PPIN [3]. BN is a directed acyclic graph representing casual relations among interacting variables at the nodes. Pathway models have natural representations as BN.

GRN is a model based on mRNA abundance, measured usually by microarrays, rendering an effective network of gene to gene interactions. DNA hybridization arrays simultaneously measure the expression levels of thousands of genes. Clustering-based

J.C. Rajapakse, B. Schmidt, and G. Volkert (Eds.): PRIB 2007, LNBI 4774, pp. 300–310, 2007.

visual tools, such as hierarchical clustering [7] and SOM [8] assume that each gene belongs to only one cluster. Such algorithms attempt to locate groups of genes having similar expression patterns over a set of experiments and hence possibly co-regulated or having similar functions. This assumption fails where genes belong to two or more independent expression patterns. Traditional statistical methods for computing low-dimensional or hidden representations of these data sets, such as principal component analysis (PCA)[9] and independent component analysis (ICA)[10], ignore the underlying interactions and provide a decomposition based purely on *a priori* statistical constraints on the computed component signals.

Here our knowledge about a biological system is not directly expressed by a parameter vector of state variables, but instead is about the statistical dependencies (or independencies) called casual relationships among the variables. The casual dependencies among variables are represented by BN in terms of conditional probabilities, so they infer 'cause and effect' relationships. The nodes of BN mimicking GRN represent gene expressions, either by analog or discrete variables, and interactions by discrete and continuous multidimensional distributions [4]. Further, dynamic Bayesian networks (DBN) can model the stochastic evolution of a set of random genes over time and therefore temporal information of interactions efficiently [5]. DBN have advantages over hidden Markov models (HMM) whose parameterization grows exponentially with the number of state variables and over Kalman filters which is capable of handling only unimodal posterior distributions. BN and DBN are defined by a graphical structure and a set of parameters, which together specify a joint distribution over the variables it represents. The nodes in Bayesian network could represent either binary or continuous variables. One advantage of representing state variables as continuous Gaussian rather than discrete is that the posterior can be marginalized efficiently over time [6]. A special class of regulatory network models is one of linear time continuous models [11]. Analysis of gene expression reveals a considerable amount of time delayed interactions, suggesting that time delay is ubiquitous in gene regulation. State-space models with time delays of gene regulatory networks use Boolean variables to capture the existence of discrete time delays of the regulatory relationships among the internal variables [12].

Various tools are now available to generate GRN from Microarray data using above models. *Gene Networks* [13] offers four models including the linear model, and 3 genetic algorithm based models, S-system, Boolean networks, and Bayesian networks. BN uses a genetic algorithm adapted from REVEAL[14] to optimize the cost function which is a NP-hard problem. Linear differential model assumes that the change of each component over time is given by a weighted sum of all other components. In this model, the expression state at one time point determines the expression state observed at the next point However assumption of linear gene-regulation relationship in unrealistic, complex systems, such as gene expression networks and metabolic pathways, are comprised of numerous richly interacting components. By representing states as binary variables and then connections by multinomial distributions, non-linear interactions among nodes can be represented in Bayesian networks.

The GRN derived from gene expression data are often over-fitted. And some of the genes are masked by the activation of highly expressed similar genes. Here we try to enhance and validate GRN derived using Bayesian networks with corresponding

PPIN discovered from PPI databases. Validation of GRN is of vital importance for making inference on large scale pathways. Here we assume skipping of one or more genes in predicted gene interaction networks and, when mapping to a protein-protein interaction, allow for prodigies of genes. As seen later, this enhances the accuracy of GRN derived from gene expression data and increases true prediction of interactions without altering biological pathways.

We demonstrate our technique with the yeast cell-cycle data, which contain differentially expressed genes in different phases of cell-cycle. Our results show that the sensitivity of BN in detecting genes of a common pathway can be improved with the validation using PPI. This paper is organized as follows: in Section 2, we explain how GRN are derived using BN Section 3 describes how GRN and PPIN are mapped. Experiments and results with yeast cell-cycle data are given in Section 4. Lastly, we draw conclusions from our findings.

2 Gene Regulatory Networks

2.1 Dynamic Bayesian Networks

A BN is a graphical model representing joint multivariate probability distributions to capture the properties of conditional independencies among variables and consists of two components: a directed acyclic graph (DAG) structure, S, and a set of conditional distributions with parameters θ, of each variable, given its parents [15]. BN are unable to model stochastic systems evolving over time. Furthermore, they are unable to construct cyclic regulations (positive and negative feedback loop mechanisms) to regulate the activities of state variables at nodes typical of biological processes. Hence, we use dynamic Bayesian networks (DBN) to generate GRN. DBN makes the following assumptions: (1) the genetic regulation process is first-order Markovian, i.e., the expression state of one gene at one time point is dependent only on the expression state of other genes observed at the previous time point; (2) the dynamic casual relationships between genes are invariable over all the time slices, that is, the set of variables and probability definitions of a DBN are the same for each time points (i.e., stationarity).

The dynamics of the DBN are hence defined in a *transition* network over two time slices, taken at time t and time $t+1$ as illustrated in Figure 1: The parameters are the probabilities of each variable, conditioned on the other variables at the previous one time point. Given the transition network over two time slices, the DBN is obtained by unrolling static transition BN over all time instances to determine the dynamics of stochastic variables over entire experiment.

In a GRN, the nodes of the BN are represented by the expressions of genes and the edges by the causal effects. Let us consider a Bayesian network representing a set of gene expressions $X = \{X_1, X_2, ...X_n\}$ in a GRN consisting of n genes. The joint probability of the expression of the genes is then be represented by $P(X) = \prod_{i=1}^{n} P(X_i \mid \prod_i)$ where $\prod_i$ denotes the set of gene expressions of parent nodes of gene i with expression X. We see that this metric is NP-hard but decomposable.

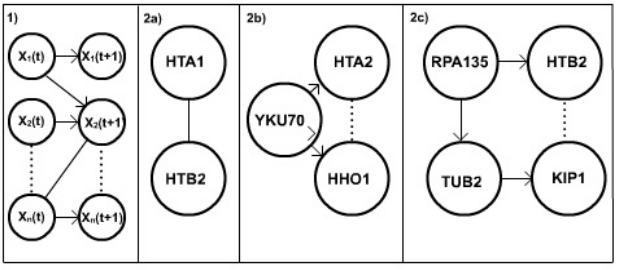

Fig. 1. Illustration of the transition network defining a dynamic Bayesian network consisting of *n* nodes. **Fig. 2.** Dotted line show the predicted interaction by a GRN (a) HTA1 interaction with HTB2 predicted by PPI which is same as by GRN (0-skip model) (b) HTA2 interaction with HHO1 predicted by PPI (1-skip model), GRN skips the gene YKU70 (c) HTB2 interaction with KIP1 predicted by PPI (2-skip model), GRN skipped 2 genes RPA135 and TUB2.

Finding a Bayesian network that fits the gene expressions best requires a search over the model space of both structure S and the interactions. Hence, a proper scoring function is needed to rank possible solutions and find the optimal solution. The posterior probability of a GRN, S, given gene expression data X, is given by, $P(S \mid X) \propto P(X \mid S)P(S)$ where $P(S)$ gives the prior probability of the network structure and $P(X \mid S)$ the likelihood. We have taken 6 important assumptions. Firstly we assume a multinomial sample, given domain U and database X, let X_l denote the first $l-1$ cases in the database. In addition, let x_{il} and $\prod_{il}$ denote the variable x_i and the parent set $\prod_i$ in the 1^{th} case, respectively. Then for all network structures B_s in U, there exist positive parameters Θ_{B_s} such that, for $i = 1,...,n$ and for all $k, k_1,....,k_{i-1}$, $p(x_{il} = k \mid x_{1l} = k_1,...., x_{(i-1)l} = k_{i-1}, X_l, \Theta_{B_s}, B_s^h, \xi) = \theta_{ijk}$. Where ξ is the current state of information. Second assumption is of parameter independence, given network structure B_s if $p(B_s^h \mid \xi) > 0$ then $\rho(\Theta_{B_s} \mid B_s^h, \xi) = \prod_{i=1}^{n} \rho(\Theta_i \mid B_s^h, \xi)$, for $i = 1,...., n : \rho(\Theta_i \mid B_s^h, \xi) = \prod_{j=1}^{q_i} \rho(\Theta_{ij} \mid B_s^h, \xi)$. Third assumption is that of parameter modularity which says that given two network structures B_{s1} and B_{s2} such that $p(B_{s1}^h \mid \xi) > 0$ and $p(B_{s2}^h \mid \xi) > 0$, and x_i has the same parameters in B_{s1} and B_{s2}, then $\rho(\Theta_{ij} \mid B_{s1}^h, \xi) = \rho(\Theta_{ij} \mid B_{s2}^h, \xi)$ $j = 1,...., q_i$. Fourth is the assumption that the distribution is Dirichlet. Given the network structure B_s such that

$p(B_s^h \mid \xi) > 0$. $\rho(\Theta_{ij} \mid B_s^h, \xi)$ is Dirichlet for all $\Theta_{ij} \subseteq \Theta_{B_s}$. That is, there exists exponents $N_{ijk}^{'}$ which depend on B_s^h and ξ, that satisfy $\rho(\Theta_{ij} \mid B_s^h, \xi) = c.\prod_k \theta_{ijk}^{N_{ijk}^{'}-1}$ where c is a normalization constant. The fifth assumption is that the database is complete. That is there are no missing data. The final assumption is of likelihood equivalence that given two network structures B_{s1} and B_{s2} such that $p(B_{s1}^h \mid \xi) > 0$ and $p(B_{s2}^h \mid \xi) > 0$, if B_{s1} and B_{s2} are equivalent, then $\rho(\Theta_U \mid B_{s1}^h, \xi) = \rho(\Theta_U \mid B_{s2}^h, \xi)$. The assumption of likelihood equivalence when combined with the previous assumptions introduces constraints on the Dirichlet exponents $N_{ijk}^{/}$. The result is a likelihood-equivalent specialization of the BD metric , which we call the BDe metric. The marginal likelihood can be represented by the BDe metric [16].

$$\text{BDe} = \prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma(N_{ij}^{/})}{\Gamma(N_{ij}^{/}+N_{ij})}\prod_{k=1}^{r_i}\frac{\Gamma(N_{ijk}^{/}+N_{ijk})}{\Gamma(N_{ijk}^{/})}$$

where $\Gamma(x)$ is a Gamma function Dirichlet distribution. Each gene i can take a finite number of distinct states *r* such that $X_i = \{x_1, x_2, ...x_{r_i}\}$ and is assumed to have a finite number of distinct state combinations of the parents, q_i such that $\prod_i = \{a_1, a_2, ...a_{q_i}\}$. $N_{ijk}^{/}$ represents the Dirichlet prior parameters and N_{ijk} the counts of interactions.

2.2 Derivation of GRN Using a Genetic Algorithm

A Genetic Algorithm (GA) is applied to effectively search the large solution space and to learn the network structure optimizing the BDe metric. We only consider binary interactions and therefore a solution individual is represented as a binary matrix which indicates the interaction states between genes and their parent genes (the genes that regulate them) where 1 denotes a regulation and 0 means no interaction. The solution $C = \{c_{i,j}\}_{n \times n}$, where $c_{ij} \in \{0,1\}$ is the interaction between genes i and j. Using the solution C we can calculate the terms $N_{ijk}^{/}$, the parameters of prior [17], and N_{ijk}, the number of observations (for the state defined by i , j and k) respectively where $X_i = x_k$, hence k is state of gene i , also $\prod_i = a_j$, j is the state combination of parents of *i*. Further, $N_{ij}^{/} = \sum_{k=1}^{r_j} N_{ijk}^{/}$ and $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. Then, using the equation above, we can get the BDe metric of the solution C.

The inputs to the genetic algorithm is a time-series data of expression of all genes. Genes in consecutive time points having similar expression levels can be said to have an interaction. The algorithm is as follows:

Procedure for DBN-GA
Begin
Initialize: Randomly create P initial individuals that can be represented as a binary interaction matrix.
While(until G generations)
Evaluate the fitness function of each individual using BDe metric
Select the elite individual to be passed on to next generation
Generate new individuals by selection, crossover and mutation. With the exception of the elite individual, the design code of each child (new individual) is created based on the design codes of two parents (old individual). Two parents are selected from the P individuals according to the probability proportional to their order of fitness (ranking or roulette strategy).
End of While
Build the gene regulation matrix based on the individual that has the largest fitness.
End

2.3 Missing Data

A key problem for all models is a shortage of data. The raw gene expression data, usually in the form of large matrix, may contain missing values. This is a result of insufficient resolution, image corruption, or simply due to dust or scratches on the slide. *KNNimpute* (K Nearest Neighbors) method [18] is used to predict missing Microarray expression levels.

3 Mapping of GRN and PPI

3.1 Protein-Protein Interaction Networks (PPIN)

Proteins frequently bind together in pairs or larger complexes to take part in biological processes. Most biological phenomena is due to a protein-protein interaction. There are several experimental techniques for determining protein-protein interaction data. Synthetic lethality[19], Affinity Capture-MS[20], and Yeast-2-Hybrid [21]being the top few in our biogrid dataset.

3.2 Motivation

The derivation of BN, using the GA, is very sensitive to the population set of structures. Since we are trying to achieve a final maximum fitness, it is at the expenses of finding the set of solutions that are together most likely to be correct, which means individual correct solutions are left out because of this evolutionary population model. The networks or the solutions on the other hand aim to connect

genes which have similar expression profiles. Since a child gene follow the expression pattern of a parent gene which is regulating it. This results in skipping or missing genes in GRN, especially those with highly expressed genes. Therefore, often the GRNs derived using BN are often underestimated in the number of genes. In order to overcome this, we propose a technique that incorporates the knowledge from corresponding PPIN to infer the missing interactions in GRN.

3.3 K-Skip Validation

In order to account for the missing genes and interactions in GRN, we employ *k*-skip models of GRN which assumes that *k*-genes are skipped in estimating GRN between two parent genes. The simplest is called the *one-skip* model where one gene is skipped in GRN due to an interaction between two genes. One reason for this could be that mRNA from gene1 might not be directly interacting with mRNA from gene2. Rather the protein product from gene1 may alter the level of mRNA from gene2. An example could be a transcription factor, which may not occur by making more of it, but just by phosphorylation (post-translational modification) [22]. Also we are interested in finding genes which lie in the same pathway. Hence these one-skip and two-skip predictions are also of high importance to us.

These models are defined as follows:

***0-skip Model*:** Indicates a direct interaction between proteins A and B

***1-skip Model*:** There exists a protein C such that both A and B interact with C according to 0-skip Model

***2-skip Model*:** There exists a protein D such that D interacts with A by 0-skip model and B by 1-skip model or vice versa.

***3-skip Model*:** There exists a protein D such that D interacts with A by 0-skip model and B by 2-skip model or vice versa.

We illustrate the above different models in the Figure 2. Figure 2 (a) shows a Gene Interaction predicted : HTA1-HTB2, which has a corresponding interaction in PPI db. This will lie in the 0-skip model. (b) Shows an interaction HHT1-HTB2 which is not found in the PPI db, however a missing gene HTA1, shows they lie in the same pathway. This is called the 1-skip model. Similarly, (d) is an example of 2-skip model. We run BN on each of the 4 sets of genes under different values of two parameters namely, the number of generations and number of individuals in each generation (i.e. population size) at the genetic algorithm step. It is possible that the interaction incorrectly bypassed a single or multiple genes. The Gene Network software provides us with the Regulatory Matrix of the final optimal solution C.

4 Cell-Cycle Regulation

4.1 Data

We illustrate our method using an application to cell-cycle regulation in yeast. Yeast has 40% genes have orthologus to human. Also it is non-pathogenic and hence can be tested for different interactions safely. We model GRN of the genes involved in the cell-cycle from an extended Spellman yeast dataset, which consists of mRNA

measurement of 6,178 genes of yeast *S. cerevisiae* [24]. Here we use the cdc15 experimental data where cdc15 yeast strain is given a cdc-15 arrest (to the cell-cycle) by moving into an incubator at 37°C. The arrest is then removed by moving back to 23°C. Cells are then monitored together at different time points for presence of new buds. 24 such time points are available from 10 to 290 mins. Cell-cycle control of transcription seems to be a universal feature of proliferating cells. Three main transcriptional waves which roughly coincide with three main cell-cycle transitions: initiation of DNA replication, entry into mitosis, and exit from mitosis. Proliferation of all cells is mediated though cell-division cycle which consists of four main phases: genome duplication (S phase) and nuclear division (mitosis or M phase), separated by two gap phases (GI and G2). Transcription of a number of genes peaks at specific cell-cycle phases. At the end of G1 phase, cells decide whether to commit to cell division in a process called start in yeast or restriction point in mammalian cells [23]. In this paper, we attempt to demonstrate our method by modeling GRNs involved in different phases of yeast cell cycle and then validating with the use of PPI data.

We downloaded the list of phase specific genes from [24]. Our dataset consists of 118 genes in G1, 36 genes in S phase, 34 genes in G2 and 60 genes in M phase respectively. Figure 3 shows the expression patterns of the 4 sets. We can see that G1 genes peak in time points 10 to 70 mins, then the S phase genes peak from 30 to 90mins, next is the G2 phase peaking 70 to 100 mins and lastly the M phase genes from time points 90 to 130 mins. Hence we can say that they are all differentially expressed.

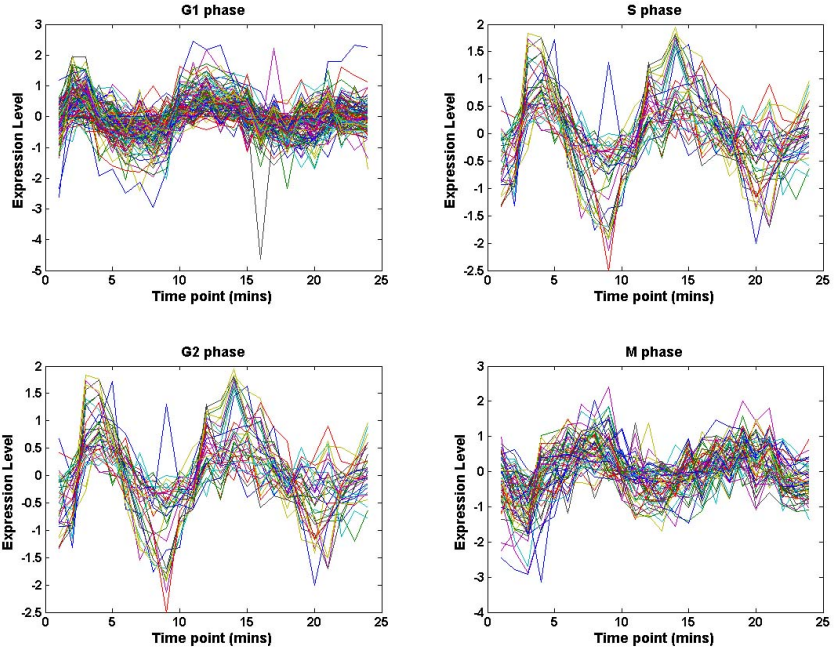

Fig. 3. Expression levels of genes in different phases of yeast cell-cycle measured at 24 time points in the cdc-15 experiment

4.2 Experiments and Results

Sensitivity of the model is very low. One of the methodologies proposed by us to overcome this is the *K*-skip model. The GRN software allows us to choose the number of Generations and population size of each generation, allowing for choice of combinations for tuning the correct number of predictions. Bayesian nets on our 4 gene sets of cell-cycle under five experimental settings is presented (Figure 4). These predictions were then validated against the PPI data for inferring the correct predictions under the *k*-skip model. As seen the accuracy of DBN first increase and then decreases with the increase in complexity of searches.

The cumulative curves for the correct number of predictions for four datasets is shown in Figure 4, G1 phase, S phase, G2 phase and M phase. We downloaded yeast data from BIOGRID [25] and got a non-redundant validations dataset of 53,235 protein interactions. It is observed that in all the graphs, there is a steep increase in the number of predictions by the one-skip model. Further increase is seen with the two-skip model. However the three-skip model shows 0 interactions in all datasets. Hence while reading Bayesian nets one must take into account that the predictions might be bypassing one or two genes in the pathways.

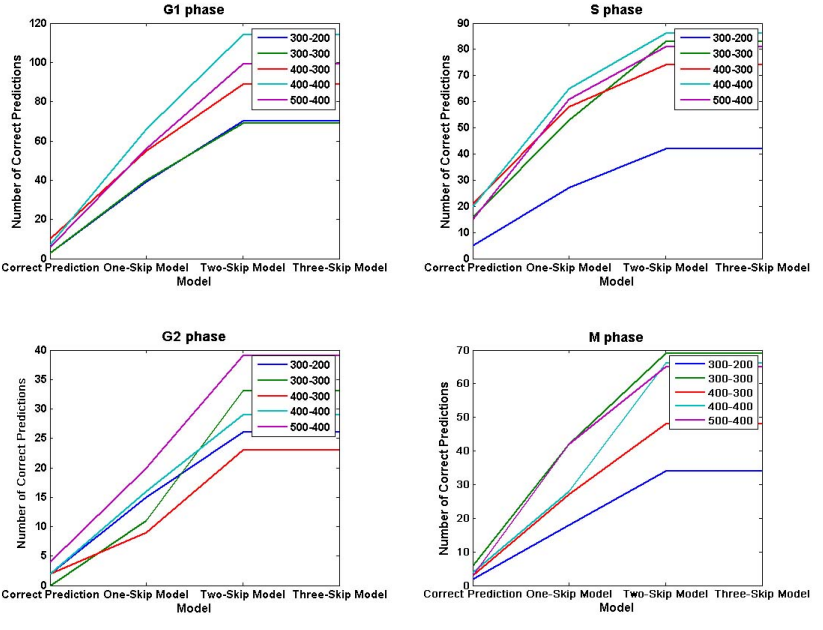

Fig. 4. Cumulative number of correct predictions for correct, one-skip, two-skip and three-skip model under 5 parameter settings (i) 300 Generations, 200 Individuals, (ii) 300 Generations, 300 Individuals, (iii) 400 Generations, 300 Individuals, (iv) 400 Generations, 400 Individuals, (v) 500 Generations, 400 Individuals

Table 1. Precision of Bayesian Networks for different cell-cycle phases. Average Precision is calculated over 5 runs with different parameter settings.

	Number of Genes	Average Precision	Maximum Precision
G1	118	0.76	0.82
S	36	0.89	0.93
G2	34	0.89	0.95
M	60	0.80	0.84

Precision for each GRN for different phases of cell-cycle which is defined as True Positive / Total Linkages was calculated. (Table 1). We notice a very high precision of over 80% in most trials. Which indicates that the Bayesian network is indeed picking up most interactions, however the accuracy is constrained by the one-skip/two-skip model.

Thus the advantages of Dynamic Bayesian Network include the ability to model stochasticity, to incorporate prior knowledge, and to handle hidden variables and missing data in a principled way. However, the discretization of gene expression by Bayesian network can lead to information loss. Also determining optimal structure of Bayesian networks is an NP-hard problem. Domain experts like the readability of trees in Bayesian networks however this is at the cost of accuracy.

5 Conclusion

We see a similar trend in all the 4 phases, confirming that a one-skip or two-skip bias exists in the model. This seems like a limitation of the model, as it looks for the best possible pathway. The proposed method may have diverse applications in understanding pathways involved in diseases.

However we must realize the constraints of the model. Some genes are redundant in different stages of the cell-cycle. This can alter the graph. Also we know that protein interactions can be stable or transient . Transient interactions are on/off and require a set of conditions that promote them. Finally we are testing the accuracy of a GN against a PPI database which is mostly generated from scientific literature and is not completely experimentally verified. Future work would involve accuracy testing against previous methods and other databases.

References

1. Nikolsky, Y.: Biological networks and analysis of experimental data in drug discovery. Drug discovery today 10(9), 653 (2005)
2. Wyrick, J.J., Young, R.A.: Deciphering gene expression regulatory networks. Curr. Opin. Genet. Dev. 12(2), 130–136 (2002)
3. Fromont-Racine, M., Rain, J.C., Legrain, P.: Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. Nature Genetics 16, 277–282 (1997)
4. Duda, R.O.: Pattern classification [Book]
5. Friedman, N., Linial, M., Nachman, I., Pe'er, D.: Using bayesian networks to analyze expression data. J. Computational Biology 7(3), 601–620 (2000)

6. Lauritzen, S.: Graphical Models. Oxford University Press, Oxford (1996)
7. Luo, F., Tang, K., Khan, L.: Hierarchical clustering of gene expression data. Bioinformatics and Bioengineering, 2003. In: Proceedings. Third IEEE Symposium, 10-12 March, pp. 328–335 (2003)
8. Wei, W., Xin, L., Min, X., Jinrong, P., Setiono, R.: A hybrid SOM-SVM method for analyzing zebra fish gene expression. Pattern Recognition, 2004. In: ICPR 2004. Proceedings of the 17th International Conference, vol. 2, pp. 323–326 (2004)
9. Alter, O., Brown, P.O., Botstein, D.: Singular value decomposition for genome-wide expression data processing and modeling. Proc. Natl. Acad. Sci. USA 97, 10101–10106 (2000)
10. Lee, S., Batzoglou, S.: Application of Independent component analysis to microarrays. Genome Biology 4, R76 (2003)
11. Bar-Joseph, Z.: Analyzing Time Series Gene Expression Data. Bioinformatics 20(16), 2493–2503 (2004)
12. Fang-Xiang, et al.: A Genetic Algorithm for Inferring Time Delays in Gene Regulatory Networks, CSB (2004)
13. Wu, C.C., Huang, H., Juan, H., Chen, S.: Gene Networks: an interactive tool for reconstruction of genetic networks from microarray data. Bioinformatics advanced access (2004)
14. Liang, S., Fuhrman, S., Somogyi, R.: REVEAL: a general reverse engineering algorithm for inference of genetic network architectures. In: Pacific Symposium on Biocomputing, vol. 3, pp. 18–29 (1998)
15. Friedmann, N., Linial, M., Nachman, I., Pe'er, D.: Using Bayesian Networks to Analyze expression data. Journal of Computational Biology 7, 601–620 (2000)
16. Heckerman, D., Geiger, D., Chickering, D.M.: Learning Bayesian networks: The combination of knowledge and statistical data. Machine Learning 9, 309–347 (1999)
17. Friedman, N., Murphy, K., Russell, S.: Learning the structure of dynamic probabilistic networks. In: Proc. Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI '98), pp. 139–147 (1998)
18. Troyanskaya, O., et al.: Missing value estimation methods for DNA microarray. Bioinformatics 17(6), 520–525 (2001)
19. Davierwala, A.P., et al.: Synthetic genetic interaction spectrum of essential genes. Nature Genetics 37(10), 1147–1152 (2005)
20. Krogan, N.J., et al.: High-definition macromolecular composition of yeast RNA-processing complexes. Mol. Cell 13(2), 225–239 (2004)
21. Young, K.H.: Yeast Two-Hybrid: So Many Interactions (in) So Little Time. Biology of reproduction 58, 302–311 (1998)
22. Page, D., Ong, I.M.: Experimental design of time series data for learning from dynamic Bayesian networks. PSB 11, 267–278 (2006)
23. Bahler, J.: Cell-Cycle Control of Gene Expression in Budding and Fission Yeast. Annu. Rev. Genet. 39, 69–94 (2005)
24. Spellman, et al.: Comprehensive Identification of Cell Cycle–regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization. Molecular Biology of the Cell 9, 3273–3297 (1998)
25. Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., Tyers, M.: BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 34, D535–D539 (2006)