ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 15, March 2017

Generative model for analysing the interaction changes in the biological network

Sheryl Maria Sebastian
Department of Computer Science and
Engineering
Mar Athanasius College of Engineering
Kothamangalam, Kerala
sherylmariasebastian@gmail.com

***Abstract*— The interaction between the genes are related to the time constraints. Now the methods are not preferring the network constraints that connect the genes. Nowadays the interactions are changing with the time. So there is systemized method for relating the time changes with the gene expression during the interaction between the genes. Introducing a method for the interaction dynamics in the biological network is represented as a tractable model rooted in Markov dynamics, which is used for analyzing the dynamics of the protein interactions. Here considering the interaction as random values. This method is carried out by small set of dataset related to the interactions. This approach is manageable in linear time. This method is effectively helpful in detecting time related interactions, strength between the interactions, functional of the active interactions. For our project datasets from Saccharomyces cerevisiae were downloaded from Gene Expression Omnibus. The first dataset contains the expression of the genes in S. cerevisiae in the availability of glucose. And second dataset contains the expression of genes in S.cerevisiae during the glucose starvation Strength between the interactions has founded and plotted for genes interactions.**

***Index Terms*— Gene Expression, generative model, interaction strength, genetic perturbation.**

I. Introduction

Biological networks are changing every time. In order to know the change of the networks, a large effort has been taken for measuring the changes of gene. This data allows to find out the genes or proteins that changed with time and their connection to other cellular components. Mathematical models can help in capturing the changes in the biological network. Conventional methods of time series cannot be applied to this problem due to the small number of observations from different time points are available relative to variables. There is a risk that many genes are having similar expression on a random scan. Finding these problems, it is recently that methods are being developed to capture time related changes. Today most of the methods[6] are not considering the interaction between the genes and also dependency between the time. So these are not perfect for the problem. In this paper, we are consider the problem of identifying time related changes in the interactions in a biological network with a defined topology from time related profiles of gene data. Saccharomyces cerevisiae is the best developed protein interaction network with a high confidence[7].So here in this paper we use saccharomyces cerevisae to conduct the studies and check our models. According to MIPS [8] protein in yeast are grouped to their biological function. 0

This annotation scheme provides functional description of the proteins in a hierarchical structure to a high degree of resolution. This allows the possibility to relate functional categorisation of the components with the time related interactions between them. This reasoning leads to two important questions: (i) can we distill observations about temporal characteristics of a group of functionally similar genes?

(ii) Is it be possible to model the effect of a gene deletion or addition comparing time related interactions between the reference strain and its perturbed mutant?

Our method which uses markovian dynamics and makes a comparison with the gene insertion and deletion on the changes in the biological networks. A fundamental fact of the model is that the evolution of the interaction strengths can be modeled in terms of the functional categories of the interacting genes. To handle low size of the sample we use Bayesian approach by using appropriate parameters for the evolution of gene interactions. Information from multiple samples which differ from the reference strain in their network topology is incorporated by assuming that interactions closer to the gene deletions are affected more strongly than those further away.

II. PROPOSED METHODOLOGY

In this section the data set used, as well as the methods for gene interaction is introduced.

a) DATA SET AND INTERACTION NETWORK

Dynamic gene datasets from Saccharomyces cerevisiae were downloaded from Gene Expression Omnibus using accession numbers GSE21988 and GSE9644.The dataset which uses the saccharomyces cerevisiae which contains gradual increment in glucose availability. And this which experience change from glucose starvation to nitrogen starvation. Eight time points are being measured from the data. The other dataset contained time related gene expression profiles in SFP1 deletion mutant and its reference at six time points after measuring steadily growing cells with glucose [11]. REF is the referred strain. MUT which is the strain in which SFP1 was deleted. The yeast interaction network was built using previously published data [14-19] as well as data downloaded from BIND [20], MIPS [8], MINT [21], DIP [22] and BioGRID [23]. We compiled all the interactions in the database and retained interactions that were backed in sources, results in a high-confidence protein interaction network. We excluded protein-DNA interactions as it is the result of this interaction in the gene expression and these interactions result in a cyclic relationship between the interactions and gene expression.

b) MODEL DESCRIPTION

We represent the gene interactions as the graph g which contains edges and vertices represented as graph G = (V,E). Here in this we are taking there are S strains which is from {1….S}.each having some genetic deletion or addition when compared to the network. Under different conditions, it will be having different outputs. That is some of the edges are switched on or off. That is we can take as some edge may be active at a certain strain and may not be in others.

i) MIXTURE MODEL

A mixture model is used which is suitable to handle multiple categorieas in the procedure since the genes belong to multiple categories. This makes us to find out the relationship between functional categories and the time related evolution characteristics of the genes which comes under in the same category.

ii).ANALYSIS OF PERTURBED NETWORK

We consider the problem of multiple strains which are just slightly altered versions of the networks where a few genes have been knocked out of the network. Therefore, most of the network remains the same across strains with only the “close” neighbourhood of the knocked out genes being affected. Gene Interaction network helps us to identify the interaction level between the genes in the network. But we have in our hand DNA microarray data which allows to simultaneously monitor the expression pattern of thousands of genes. We then try to find the interaction level between the genes from DNA microarray of dataset by using the following steps:

A. Dimension reduction based on relative entropy

It is usually occurring that the number of genes which will be more than the samples so we should reduce it. First we will reduce the dimension.We aimed at ranking the genes in order of minimum relative entropy.Then select top n genes (n depends on the choice of the user) according to their ranks. This will be taken as the input for the coming steps.

B. Using k-means and C-Means method clustering the gene dataset

Clustering the dataset means grouping the data points into different groups which is having same charecteristics but if it is having large number of features then it will be more difficult. After selecting the top n genes by KL divergence we apply Principal component analysis on dataset to select the features.The genes sharing same properties fall in the same cluster, which signifies these genes are co-expressed for this particular type of cancer. This is the basic model.Some of the genes do not fall in the any cluster these are known as outliers. To determine the membership we use fuzzy c means. These genes connect the other clusters.

C. Finding the degree of interaction between genes using Correlation Coefficiency and Entropy

After performing the clustering, we will obtain a graph with unlabeled edges. To calculate the edge weight we take the cluster of each cluster which is a graph.To calculate the edge weights should calculate the interaction between the genes. Firstgenes will be filtered corresponding to the expression profiles less than 10th percentile. The outputs , represent the genes with a variance greater than a threshold with a value of 1 and the genes with a variance less than a threshold with a value of 0. This implies that a cluster having a single gene with a value of 0 does not create an edge with the other genes. However, clusters having more than one gene having a value of 0 have edges between them. The gene-gene interaction level is then estimated by calculating the correlation coefficient for each pair of genes in the same cluster.

We investigated two schemes: a distance dependent b such that starting from the deleted node, all nodes at distance (in a breadth first search from the deleted node), greater than a selected distance have b = 0. We also investigated b in the range [0, 1]. [6] discussed about a Secure system to Anonymous Blacklisting. The secure system adds a layer of accountability to any publicly known anonymizing network is proposed. Servers can blacklist misbehaving users while maintaining their privacy and this system shows that how these properties can be attained in a way that is practical, efficient, and sensitive to the needs of both users and services. This work will increase the mainstream acceptance of anonymizing networks such as Tor, which has, thus far, been completely blocked by several services because of users who abuse their anonymity.

d) ARCHITECTURE

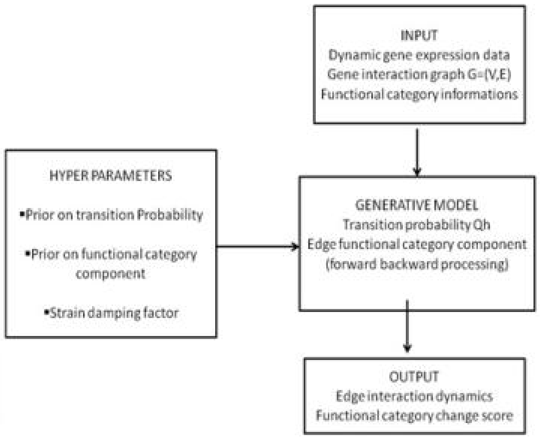
INPUT
Dynamic gene expression data
Gene interaction graph G=(V,E)
Functional category informations
HYPER PARAMETERS
•Prior on transition Probability
•Prior on functional category component
•Strain damping factor
GENERATIVE MODEL
Transition probability Qh
Edge functional category component
(forward backward processing)
OUTPUT
Edge interaction dynamics
Functional category change score

Figure 1:architecture of proposed system

III. EXPERIMENTAL RESULTS

We construct a synthetic graph G = (V, E) using the Erdos-Renyi model consisting of |V | = 1000 nodes, which represents the genes, and |E| = 5961 randomly generated edges, which represent the gene interactions. Each edge e Î E can have one of the following states .W = {−2,−1, 0, 1, 2}, signifying the interaction strength. We consider two models for interaction dynamics in the genetic network to investigate the impact of functional category information on the interaction dynamics in the gene interaction network, which are described below:

The first model incorporates the functional categories, which impact the evolution of interaction strengths as specified in eqn. (4) and eqn. (5). We use H = 200 functional categories which govern the behaviour of the evolution of the gene expression. Nodes were randomly assigned to (multiple) functional categories reflecting the empirical distribution of functional categories for genes in MIPS database [8].

The second model assumes the evolution of the interaction strengths for an edge is independent of other edges. The transition probability matrix Q_e for each edge e Î E is generated by sampling with equal probability from the two classes H0 : tr(Q_e) >0.5W and H1 : tr(Q_e) ≤ 0.5W.

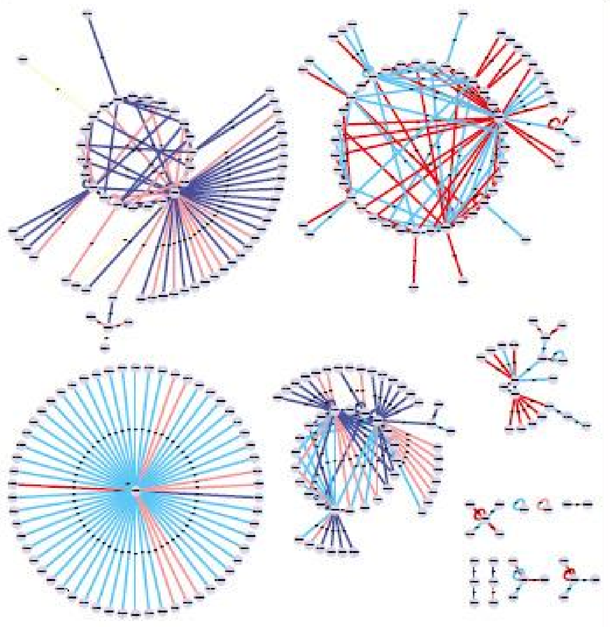

Figure 3 Experiment 1 network interactions. Time varying interaction strengths between the genes from Experiment 1. Each network is composed of all the genes from the eight clusters previously identified [11], and is shown for *t* = 0.0 hrs. The edge colors denote their interaction strength, which was classified as strong repressing (red), low repressing (pink), no effect (yellow), low inducing (light blue) and strong inducing (dark blue).

IV. CONCLUSION

This is a systematic model that relates temporal changes in gene expression data to the dynamics of interactions in the context of a regulatory network.. The framework of the model will also inherently facilitate analyzing the effect of a perturbation in the network. For a given regulatory network and gene expression data, this was able to identify time sensitive interactions in the network and determine their strength. It was able to deduce the most active functional categories that interacted. In addition to these, this uses a damping feature that models the effect of a network perturbation by localizing more activity around the point of perturbation. These three novel features that offers reflect its advantage over many other time-series models that have been developed recently. Of particular interest is its ability to capture abrupt changes in the interaction patterns. For example, NETGEM identified momentary arrest in ribosome biosynthesis during the transition in the nutrient that limits growth from glucose to ammonia (Experiment 1). We identified many actively interacting genes that were

implicated to play an important role in the biological conditions from which we obtained the data.
This lends the promise that new insights obtained from using NETGEM are also physiologically relevant. Given that the inputs to NETGEM are the topology of the network and temporal variation of the nodes, it is evident that this methodology has widespread applications in analyzing network dynamics, beyond biological systems.

REFERENCES

[1]. Stigler B, Jarrah A, Stillman M, Laubenbacher R: Reverse engineering of dynamic networks. Annals of the New York Academy of Sciences 2007,1115:168-77.
[2]. Glass L, Kaplan D: Time series analysis of complex dynamics in physiology and medicine. Med Prog Technol 1993, 19:115-128.
[3]. Leek JT, Monsen E, Dabney AR, Storey JD: EDGE: extraction and analysis of differential gene expression. Bioinformatics 2006, 22(4):507-508.
[4]. Ernst J, Joseph ZB: STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics 2006, 7:191.
[5]. Ramoni MF, Sebastiani P, Kohane IS: Cluster analysis of gene expression dynamics. Proceedings of the National Academy of Sciences of the United States of America 2002, 99(14):9121-9126.
[6].Christo Ananth, A.Regina Mary, V.Poornima, M.Mariammal, N.Persis Saro Bell, “Secure system to Anonymous Blacklisting”, International Journal of Advanced Research in Biology, Ecology, Science and Technology (IJARBEST), Volume 1,Issue 4,July 2015,pp:6-9
[7]. Petranovic D, Vemuri GN: Impact of yeast systems biology on industrial biotechnology. J Biotechnol 2009, 144(3):204 11.
[8]. Mewes HW, Dietmann S, Frishman D, Gregory R, Mannhaupt G, Mayer KF, Munsterkotter M, Ruepp A, Spannagl M, Stumpflen V, Rattei T: MIPS: analysis and annotation of genome information in 2007. Nucleic acids
research 2007, 36.
[9]. Guo F, Hanneke S, Fu W, Xing E: Recovering temporally rewiring networks: A model-based approach. Proceedings of the 24th ICML 2007.
[10]. Song L, Kolar M, Xing EP: KELLER: estimating time-varying interactions between genes. Bioinformatics 2009, 25(12):i128-136.
[11]. Farzadfard F, Nielsen ML, Nielsen J, Vemuri GN: Metabolic and transcriptional dynamics during the transition from carbon limitation to nitrogen limitation in Saccharomyces cerevisiae. BMC Genomics (in review)
2010, X.
[12]. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G,Tierney L, Yang JY, Zhang J: Bioconductor: open software development for computational biology and bioinformatics. Genome Biology 2004, 5(10):R80.
[13]. Wu Z, Irizarry RA: Stochastic models inspired by hybridization theory for short oligonucleotide arrays. J Comput Biol 2005, 12(6):882-893.
[14]. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G: Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002, 415:141-147.
[15]. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P,
Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreault M, Muskat B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Sorensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M: Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 2002, 415:180-3.
[16]. Gavin AC, Aloy P, P Grandi RK, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C,