10

Computational Modelling Strategies for Gene Regulatory Network Reconstruction

Muhammad Shoaib Sehgal[1], Iqbal Gondal[2] and Laurence Dooley[3]

[1] Faculty of IT, Monash University, VIC 3842 Churchill Australia and Victorian Bioinformatics Consortium, Clayton VIC 3800 Australia
Shoaib.Sehgal@infotech.monash.edu.au
[2] Faculty of IT, Monash University, VIC 3842 Churchill Australia and Victorian Bioinformatics Consortium, Clayton VIC 3800 Australia
Iqbal.Gondal@infotech.monash.edu.au
[3] Faculty of IT, Monash University, VIC 3842 Churchill Australia
Laurence.Dooley@infotech.monash.edu.au

Summary: Gene Regulatory Network (GRN) modelling infers genetic interactions between different genes and other cellular components to elucidate the cellular functionality. This GRN modelling has overwhelming applications in biology starting from diagnosis through to drug target identification. Several GRN modelling methods have been proposed in the literature, and it is important to study the relative merits and demerits of each method. This chapter provides a comprehensive comparative study on GRN reconstruction algorithms. The methods discussed in this chapter are diverse and vary from simple similarity based methods to state of the art hybrid and probabilistic methods. In addition, the chapter also underpins the need of strategies which should be able to model the stochastic behavior of gene regulation in the presence of limited number of samples, noisy data, multi-collinearity for high number of genes.

Key words: Gene Regulatory Networks, Deterministic Modelling, Stochastic Modelling and Computational Intelligence Methods for GRN Modelling

10.1 Introduction

Basic cellular functionality is highly dependent on the transcriptional process of DNA to form proteins. For production of proteins, a DNA is first converted to mRNA (Transcription, Fig. 10.1) which then leads to the production of proteins (Translation, Fig. 10.1) where the basic production codes are provided by the genes for the synthesis of proteins. Several statistical and computational intelligence techniques have been used for class prediction [1–3], differentially expressed gene selection [4, 5] and to cluster functionally related genes under variety of conditions. Even though,

Muhammad Shoaib Sehgal et al.: *Computational Modelling Strategies for Gene Regulatory Network Reconstruction*, Studies in Computational Intelligence (SCI) **85**, 207–220 (2008)
www.springerlink.com

these techniques give biologists valuable insights of different biological systems but still there is a need of methods which can model uncertainty, can cope with thousands of genes at a time and help to understand complex genetic interactions. In addition to that, since most of the analysis are based on over/under expressed genes studies despite the fact that differential expression analysis doesn't harness full potential of microarray gene expression data because genes are treated independent of each other and interactions among them are not considered [6]. *Gene Regulatory Network* (GRN) can model how genes interact with each other to regulate different metabolism to carry out the cellular functionality [7]. This GRN modelling has overwhelming applications in biology starting from diagnosis through to drug target identification.

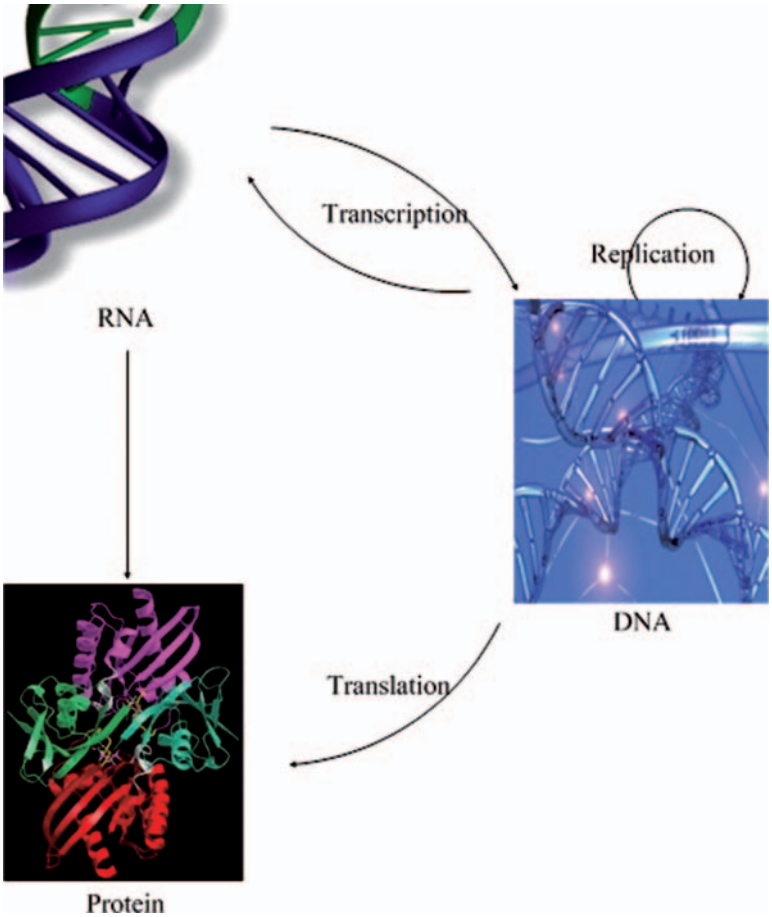

Fig. 10.1. Central Dogma [35–37]

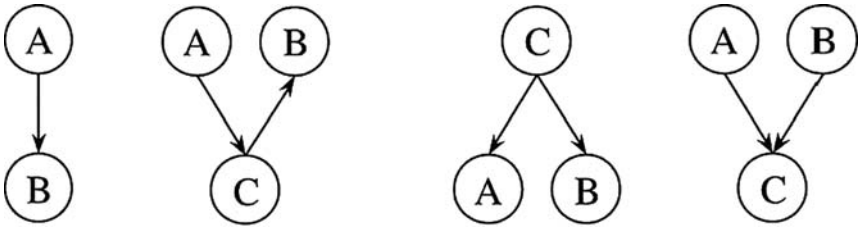

Fig. 10.2. Interaction Patterns [9]

A link between two genes g_i and g_j shows that the product of a gene g_i can inhibit or activate gene g_j which means that the protein product of gene g_i is a transcriptional factor that binds to the operator site in the promoter region of gene g_j and regulates its expression levels. Normally the links are analyzed by number of regulating *Transcriptional Factors* (TF) (incoming links to a gene) and the number of regulated genes per TF i.e. number of out coming links due to the inherent directionality property [8], to determine their distribution is followed by power law or exponent like models and to find hub genes (Hubs: The genes with max number of links). For example yeast network belongs to mix class of networks i.e. power and exponent. There are various possibilities by which genes interact with each other which are outlined in Fig. 10.1. A gene can directly trigger the other gene (Direct Link), a gene can indirectly trigger the other gene (Indirect Interaction), a gene can activate or repress two or more genes (Divergence) or two or more genes activate/repress a gene (Convergence).

Gene network construction, however, is a difficult task due to noisy nature of microarray data, curse of dimensionality (number of features are much higher than number of samples) and multi-collinearity [10]. Several techniques have been developed to model these Gene Regulatory Networks but in general GRN reconstruction consists of following series of steps (Fig. 10.1): Firstly, a sample is prepared under experimental conditions for example yeast for heat shock etc. Then microarray gene expression data is generated from the prepared sample. This is followed by a normalization step and then GRN is constructed using GRN modelling methods. The GRN modelling methods are diverse and it is important to have their in depth understanding and to know their relative strengths and weakness [9]. This chapter will provide details of commonly used Computational Intelligence methods such as: Similarity Based Methods, Probabilistic, Deterministic, Boolean and hybrid modelling techniques with their respective pros and cons.

In detailing the various GRN modelling strategies, a gene expression matrix Y is assumed to have m rows and n columns where the columns represent genes and rows represent samples as in Eq. 10.1. A gene expression vector in sample i can be referred as g_i

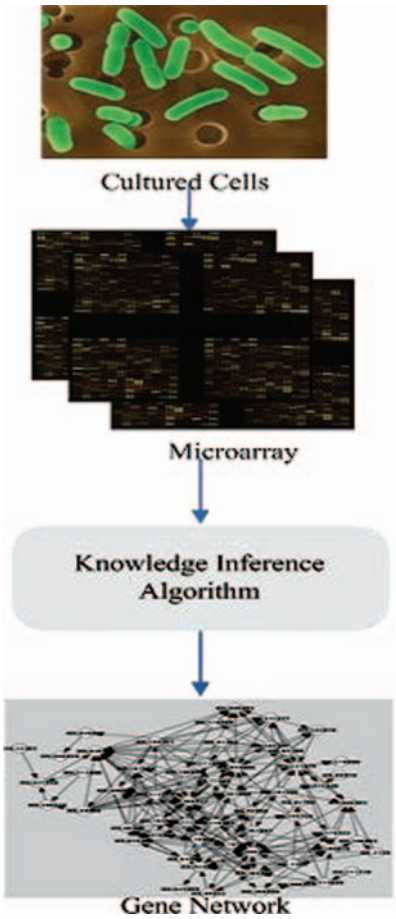

Fig. 10.3. GRN Reconstruction Framework [38, 39]

$$Y = \begin{bmatrix} g_{11} & g_{12} & g_{13} & \cdots & g_{1n} \\ g_{21} & g_{22} & g_{23} & \cdots & g_{1n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ g_{m1} & g_{m2} & g_{m3} & \cdots & g_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n} \tag{10.1}$$

10.2 Pair Wise GRN Reconstruction Methods

The regulatory interactions can be modelled by pair wise genetic interactions in which both the regulators and targets can be modelled by similarity measures. The algorithm consists of three major steps [11]:

1. Compute pair wise similarity/dissimilarity measure between each pair of genes in Y.

2. Rank all the genes based on their relative similarity/dissimilarity values.

3. Use a cut off threshold value δ to select co-regulated genes.

Following sub sections will outline some of the commonly used similarity measures and their respective merits and demerits.

Correlation and Distance Functions

Pearson correlation is the most commonly used similarity measure [12], to find the co-regulated links [13] due to its simplicity and relatively better performance for microarray data [14]. Pearson correlation r between two genes g_i and g_j can be computed as:

$$r = \frac{\sum g_i g_j - \frac{\sum g_i g_j}{n}}{\sqrt{\left(g_i^2 - \frac{(\sum g_i)^2}{n}\right)\left(g_j^2 - \frac{\left(\sum g_j\right)^2}{n}\right)}} \tag{10.2}$$

However, Pearson correlation can lead to spurious correlate genes as it uses absolute gene expression values to compute the similarity. Also, the method is highly sensitive to outliers [15] between the genes so an alternative solution is to use rank statistics like Spearman ranked correlation, as proposed in [16] or Jackknife correlation [17], though the former is computationally intensive. Spearman correlation between two genes g_i and g_j can be computed as:

$$\rho = \frac{6\sum D_g^2}{N_g\left(N_g^2 - 1\right)} \tag{10.3}$$

where D_g is the distance between ordered pairs of genes g_i and g_j and N_g is the number of pairs. Several dissimilarity measures like Euclidean Distance, Manhattan metric, percent remoteness, chord distance and geodesic distance [15] can also be applied to find gene co-regulation. While the most common one being Euclidean distance, as proposed in [18]. The Euclidean distance is simple to compute and is less computational intensive though, it is sensitive to outliers [19] and can ignore the negative correlations [3] so it can ignore the co-regulation involving repression of a gene by the other gene.

Mutual Information

The mutual information $I(g_i, g_j)$, between two discrete gene expression vectors g_i and g_j can be computed as:

$$I(g_i, G_j) = \frac{1}{m} \sum_{k=1}^{m} log \left[\frac{p(g_{ik}, g_{jk})}{p(g_{ik}) p(g_{jk})} \right] \tag{10.4}$$

where $P(g_i, g_j)$ is a joint probability, $P(g_i)$ and $P(g_j)$ are respective marginal probabilities of expression vectors and m is number of samples. Since, the gene expression values are continuous values so they are discredited prior using above definition. This process however, can lose the information [20] therefore, various methods are proposed to compute the mutual information from continuous variables with the famous one being *Gaussian Kernel Estimator* to compute the probabilities and can be expressed as:

$$p(g_{ik}) = \frac{1}{\sqrt{2\pi N \alpha_1}} \sum_{l} e^{\frac{(g_{ik} - g_{il})}{2\alpha_1^2}} \tag{10.5}$$

$$p(g_{jk}) = \frac{1}{\sqrt{2\pi N \alpha_1}} \sum_{l} e^{\frac{(g_{jk} - g_{jl})}{2\alpha_1^2}} \tag{10.6}$$

$$p(g_{ik}, g_{ik}) = \frac{1}{\sqrt{2\pi N \alpha_2}} \sum_{l} e^{\frac{(g_{ik} - g_{il}) + (g_{jk} - g_{jl})}{2\alpha_2^2}} \tag{10.7}$$

where α_1 and α_2 are tunable parameter and can be computed by Monte Carlo Simulations [21] using bi-variate normal probability densities [22].

The mutual information between two variables is always ≥ 0, where mutual information zero means two genes are functionally independent of each other. Mutual information is considered to be providing a more general framework than correlation and dissimilarity measures like Pearson correlation to measure the dependency between the variables [12] due to its theoretical and probabilistic basis.

The above mentioned similarity measures use the complete expression profiles of the data to compute the degree of similarity and normalized the data to remove the expression profiles which have insignificant changes. The transcriptional regulators which act as switches in a transcriptional network however, may be expressed at very low levels so normalization can miss such key regulations. Also, above mentioned similarity measures are normally sensitive to noise and outliers like Pearson correlation. Moreover, the relationship between the genes is often expressed local similar patterns rather than global patterns which can be missed if the complete expression profiles are considered for similarity measure [23]. Above mentioned problems can be addressed by using local shape based similarity measures which will be explained in greater detail in forthcoming Sub Section.

Local Shape Based Similarity

Local shape based similarity method introduced by Balasubramaniyan *et al.* [23] searches for local relationships between genetic expressions. The similarity between genes g_i and g_j is computed by:

$$Sim(g_i, g_j) = \max_{k_m \leq k \leq n} Sim_k(g_i, g_j) \tag{10.8}$$

where $Sim_k(g_i, g_j) = \max_{1 \leq l, o \leq n-k+1} S(g_i[l, l+k-1], g_j[o, o+k-1])$, S is a similarity measure, k is the lower bound of length of alignment and k is the best alignment length. The *Sim* is computed like sequence alignment algorithm such as, BLAST, Needleman-Wunsch [24], Smith-Waterman algorithm [25] or simple sliding window algorithm. Since Spearman correlation uses ranks to compute the similarity therefore Balasubramaniyan et al, suggested the use of Spearman correlation for their proposed local shape based similarity measure. Local shape based similarity algorithm though, claimed to be extracting locally similar patterns lacks evidence that the method can extract non linear relationships between the genes especially, when it is using Spearman correlation as a similarity metric.

10.3 Deterministic Methods for GRN Inference

Differential Equations

Different types of differential equations have been widely used to model GRN systems for example, Nonlinear Ordinary Differential Equations, Piecewise Linear Differential Equations and Qualitative Differential Equations. The ordinary differential equations model the rate of the regulation of a gene as a function of expression values of other genes by:

$$\frac{dg_{ik}}{dt} = f_k(g_i), l \leq k \leq n \tag{10.9}$$

where $g = [g_1, \cdots, g_n] \geq 0$ contains concentrations of different interacting genes, proteins or small molecules [26] where discrete time delays to model transcription, translation and diffusion can be represented as:

$$\frac{dg_{ik}}{dt} = f_k(g_1(t-\tau_{k1}), \cdots, g_n(t-\tau_{kn}))1 \leq k \leq n \tag{10.10}$$

where $\tau_{k1}, \cdots, \tau_{kn} > 0$ represents discrete time delay. The differential equations has the ability to scale up to genomic level and can incorporate the delay between transcription and translation [27]. However, due to their deterministic change assumption i.e. d/dt is not always valid due to cellular fluctuations. Also, differential equations implicitly assume that the GRN system is spatially homogeneous which is not always true [26] which can lead to erroneous inference.

Boolean GRN Modelling Methods

The level of gene expression can depend on multiple transcriptional factors and thus on many genes. So the working principal of gene co-regulation can be modelled by using Boolean network model which functionally relate expression states of genes

with the other genes using Boolean logic rules. For example, some genes are activated by different possible transcriptional factors so they can be connected using OR logic (Fig. 10.4) while other require two or more genes to be involved in a gene regulation (AND logic). Similarly inhibitory relationship between genes can be modelled by NOT logic (Fig. 10.4) and more complex rules can be modelled using combination of Boolean logics such as, if a gene is regulated only if one of its possible activator is active, while it is not repressed its one of possible inhibitors, can be modelled by OR-NOR logic (Fig. 10.4) [28].

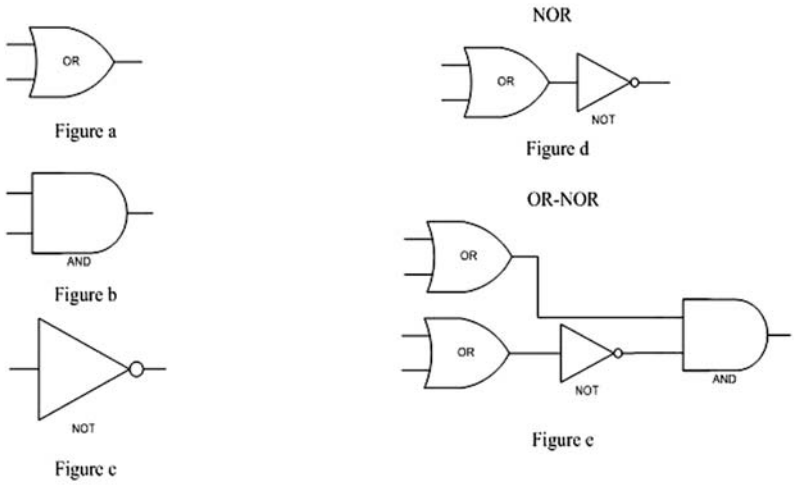

Fig. 10.4. Logical Representation of Different Gene Regulatory Relationships using Boolean Logic

10.4 Probabilistic GRN Reconstructed Strategies

Probabilistic models are one of the most commonly used GRN modelling methods due to their ability to model highly stochastic nature of gene co-regulation, as shown by most of the experimental studies [29]. The methods include: Bayesian Networks, Dynamic Bayesian Networks and various hybrid probabilistic models.

Bayesian networks model causal relationship between genes by applying probability theory [24]. A Bayesian network represents GRN by *Directed Acyclic Graph* (DAG), G(V,E) where each gene is represented by different vertex V and edge E represents the regulation pathway [20]. For instance, if gene gi is regulated by g_j it is represented by a link from g_i to g_j ($g_i \rightarrow g_j$). The dependency between a gene and its regulators is calculated by using a joint probability of a gene given its regulators.

For instance, if p_i represents the set of regulators (parents) of gene g_i then the joint probability can be computed by chain rule, such as:

$$P(g_i, g_j, \cdots, g_m) = \prod_{i=1}^{m} P(g_i|p_i) \tag{10.11}$$

Causal relationship modeling between genes using Bayesian networks can be divided into two main modules:

1. Scoring Function
2. Search Method

where scoring function computes how well the data fits to structure and search method searches for the network with the highest scores [30]. The most common scoring functions are *Bayesian Score* [31] and *Minimum Description Length* (MDL) [20, 32]. An important property of scoring functions is decomposability which can be defined as:

$$Score(S,Y) = \sum_{i} Score(g_i, P_i), Y(g_i, P_i) \tag{10.12}$$

where Score is a scoring function, $Y(g_i, P_i)$ is data involving g_i and P_i.

As alluded earlier, the second step in GRN reconstruction using Bayesian networks is defining a search function. The search problem is NP hard therefore heuristic methods are used to search the sub-optimal structure of the network. The search method applies three basic operations to the network with the objective to optimize the score: Addition, Deletion and Reversing the link direction as shown in Fig. 10.5. For each change the graph, search algorithm computes the score using scoring function and also, nullifies the invalid moves e.g. addition formed a cycle which is invalid in Bayesian networks (Fig. 10.5). Finally, the graph with the highest score is selected [33]. However it is worth noting that two graphs may have same score which is one of the disadvantages of using Bayesian networks. The most common heuristic search algorithms used in this context are: Hill Climbing, Simulated Annealing, Genetic Algorithms and K2.

The Bayesian networks takes the advantage of their sound probabilities semantics, ability to deal with noise & missing data which will help to cope with incomplete knowledge about biological system and flexibility to integrate prior biological knowledge into the system [30]. However, Bayesian networks have disadvantage of their high computational complexity, lack of scalability [17] and acyclic restriction [20]. The acyclic problem can be solved by using dynamic Bayesian networks [24] at additional computational cost by adding time delay [33].

Due to the relative advantages of above mentioned methods several hybrid methods have been evolved over the years to utilize the advantages [22] of each method.

10.5 Hybrid GRN Inference Methods

This section provides overview of different hybrid GRN reconstruction methods. Zhao *et al*. [20] introduced a hybrid model based on Mutual information and MDL

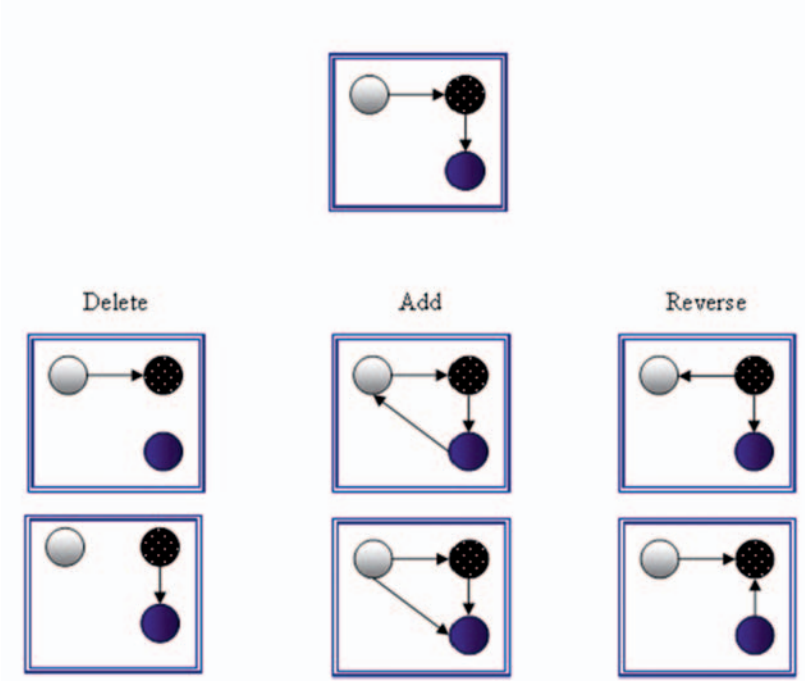

Fig. 10.5. Possible Moves of a Search Function

principal. The method computes pair wise mutual information and used MDL to automatically determine the selection threshold. It then removes the links and computes the score. Finally, the network with the highest score is selected. The method has advantage of automatically threshold selection compared to trial and error method and is scalable than simple Bayesian network. The technique however, computes discrete mutual information which can loose some valuable information.

Basso et al, introduced *Algorithm For The Reconstruction of Accurate Cellular Networks* (ARACNe) which computes the pair wise mutual information by using *Gaussian Kernel Estimator* (Section 10.2) [22]. Mutual information computation step is followed by network pruning using Data Processing Inequality which can be defined as, when two genes g_i and g_j are interacting through a third gene g_k and $I(g_i, g_j|g_k)$ is zero then these genes are directly interacting with each other if:

$$I(g_i, g_k) \leq (g_i, g_j) and I(g_i, g_k) \leq (g_j, g_k) \tag{10.13}$$

This property is asymmetric and therefore has the possibility of rejecting some of the loops or interaction between three genes whose information may not be fully modelled by pair wise mutual information. The use of tolerance threshold can solve this problem as has advantage of avoiding rejection of some of the triangular links and loops [22].

The ARACNe method is robust against noise and is proven to be showing better modelling than Bayesian networks.

Gene regulation is fuzzy in nature therefore different fuzzy GRN modelling methods have been proposed. Du *et al*. [34] introduced a method based on Fuzzy k means algorithm. The method first clusters functionally related genes using fuzzy k means algorithm and then constructs the network using following linear model:

$$g_i(i+\Gamma_i) = \sum w_{ji} g_j + b_i \tag{10.14}$$

where g_i is the expression level of i^{th} gene at time t, Γ is a regulation time delay of g_i, w_{ji} is the weight associated to the inference of g_j to g_i and b_i is the bias indicating default expression of g_i without regulation. The method finally evaluates the link strength using fuzzy metric based on Gene Ontology evidence strength and co-occurrence of similar gene functions. The method utilized fuzzy logic to model the fuzzy nature of gene co-regulation. However the method uses predetermined number of clusters which may loose some of the regulated links. To overcome this disadvantage Sehgal *et al*. [7] introduced a method, *Adaptive Fuzzy Evolutionary GRN Reconstruction* (AFEGRN) for modelling GRNs. The AFEGRN automatically determines model parameters such as, number of clusters for fuzzy c-means using fuzzy-PBM index and evolutionary *Estimation of Gaussian Distribution Algorithm*. Finally the network is reconstructed using Spearman correlation Eq. 10.3. The method adapts to the data distribution compared to the earlier described method by [34] which uses preset value of number of clusters. Since the method used Spearman correlation so it has the disadvantage like other correlation based matrices that it may introduce spurious co regulated links. To overcome this disadvantage Chen *et al*. [30] introduced a hybrid algorithm based on mutual information. The method first computes the mutual information between the genes and then uses K2 algorithm to finally construct the network. Since, the K2 algorithm is highly sensitive to missing values and microarray data contains at least 5% missing values and in most data sets, at least 60% of genes have one or more missing values [35], therefore the method can miss important regulation links.

10.6 Conclusions

Gene Regulatory Network (GRN) models the genetic interaction between the genes and other cellular components to elucidate the cellular functionality. This GRN modelling has overwhelming applications in biology starting from diagnosis through to drug target identification. Several GRN modelling methods have been proposed and it is important to study the relative merits and demerits of each method. This chapter has provided a comprehensive study on GRN reconstruction algorithms by highlighting their respective merits and demerits. The chapter introduced simple similarity based methods to state of the art hybrid and probabilistic methods. It is clear however, that despite the significant contribution of the proposed methods still there is a

need of technique which should be able to model the stochastic behavior of gene regulation in the presence of limited number of samples, noisy data, multi-collinearity for high number of genes.

Acknowledgements

The authors would like to formally acknowledge Ms. Megan Woods, Faculty of IT, Monash University, Australia for her help with the camera ready version of the Chapter.

References

[1] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasen-beek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Down-ing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lan-der, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," Science, pp. 286(5439):531-537, 1999.

[2] M. S. B. Sehgal, I. Gondal, and L. Dooley, "Statistical Neural Networks and Support Vector Machine for the Classification of Genetic Mutations in Ovarian Cancer," IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)'04, USA., pp. 140-146, 2004.

[3] M. S. B. Sehgal, I. Gondal, and L. Dooley, "Missing Values Imputation for DNA Microarray Data using Ranked Covariance Vectors," The International Journal of Hybrid Intelligent Systems (IJHIS), vol. ISSN 1448-5869, 2005.

[4] S. Dudoit, J. Fridlyand, and T. P. Speed, "Comparison of discrimination methods for the classification of tumors using gene expression data," Journal of the American Statistical Association, pp. 77-78, 2002.

[5] M. S. B. Sehgal, I. Gondal, and L. Dooley, "Collateral Missing Value Estimation: Robust missing value estimation for consequent microarray data processing," Lecture Notes in Artificial Intelligence (LNAI), Springer-Verlag, pp. 274-283, 2005.

[6] J. K. Choi, U. Yu, O. J. Yoo, and S. Kim, "Differential coexpression analysis using microarray data and its application to human cancer," Bioinformatics, vol. 21, pp. 4348-4355, December 15, 2005 2005.

[7] M. S. B. Sehgal, I. Gondal, L. Dooley, and R. Coppel, "AFEGRN- Adaptive Fuzzy Evolutionary Gene Regulatory Network Reconstruction Framework," IEEE- World Congress on Computational Intelligence-FUZZ-IEEE, pp. 1737-1741, 2006 2006.

[8] I. Farkas, C. Wu, C. Chennubhotla, I. Bahar, and Z. Oltvai, "Topological basis of signal integration in the transcriptional-regulatory network of the yeast, Saccharomyces cerevisiae," BMC Bioinformatics, vol. 7, p. 478, 2006.

[9] A. V. Werhli, M. Grzegorczyk, and D. Husmeier, "Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks 10.1093/bioinformatics/btl391," Bioinformatics, vol. 22, pp. 2523-2531, October 15, 2006 2006.

[10] G. Fort and S. Lambert-Lacroix, "Classification using partial least squares with penalized logistic regression," Bioinformatics, vol. 21, pp. 1104-1111, 2005 2005.
[11] P. Y. Chen and P. M. Popovich, Correlation: Parametric and Nonparametric Measures, 1st edition ed.: SAGE Publications, 2002.
[12] R. Steuer, J. Kurths, C. O. Daub, J. Weise, and J. Selbig, "The mutual information: Detecting and evaluating dependencies between variables 10.1093/bioinformatics/18.suppl_2.S231," Bioinformatics, vol. 18, pp. S231-240, October 1, 2002 2002.
[13] J. M. Stuart, E. Segal, D. Koller, and S. K. Kim, "A Gene-Coexpression Network for Global Discovery of Conserved Genetic Modules 10.1126/science.1087447," Science, vol. 302, pp.249-255, October 10, 2003 2003.
[14] G. Yona, W. Dirks, S. Rahman, and D. M. Lin, "Effective similarity measures for expression profiles 10.1093/bioinformatics/btl127," Bioinformatics, vol. 22, pp. 1616-1622, July 1, 2006 2006.
[15] X. Xia and Z. Xie, "AMADA: analysis of microarray data," Bioinformatics Application Note, vol. 17, pp. 569-570, 2001.
[16] M. S. B. Sehgal, I. Gondal, L. Dooley, and R. Coppel, "AFEGRN: Adaptive Fuzzy Evolutionary Gene Regulatory Network Re-construction Framework," World Congress on Computational Intelligence: Fuzzy Systems., 2006.
[17] X. J. Zhou, Ming-Chih, J. Kao, H. Huang, A. Wong, J. Nunez-Iglesias, M. Primig, O. M. Aparicio, C. E. Finch, T. E. Morgan, and W. H. Wong, "Functional annotation and network reconstruction through cross-platform integration of microarray data," Nature Biotechnology, vol. 23, pp. 238-243, 2005.
[18] L. J. Heyer, S. Kruglyak, and S. Yooseph, "Exploring Expression Data: Identification and Analysis of Coexpressed Genes 10.1101/gr.9.11.1106," Genome Res., vol. 9, pp. 1106-1115, November 1, 1999 1999.
[19] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. Altman, "Missing Value Estimation Methods for DNA Microarrays," Bioinformatics, vol. 17, pp. 520-525, 2001.
[20] W. Zhao, E. Serpedin, and E. R. Dougherty, "Inferring gene regulatory networks from time series data using the minimum description length principle," Bioinformatics, vol. 22(17), pp. 2129-2135, 2006.
[21] G. Casella and C. P. Robert, Monte Carlo Statistical Methods: Springer, 2005.
[22] K. Basso, A. A. Margolin, G. Stolovitzky, U. Klein, R. Dalla-Favera, and A. Califano, "Reverse engineering of regulatory networks in human B cells," Nature Genetics, vol. 37, pp. 382-390, 2005.
[23] R. Balasubramaniyan, E. Hullermeier, N. Weskamp, and J. Kamper, "Clustering of gene expression data using a local shape-based similarity measure 10.1093/bioinformatics/bti095," Bioinformatics, vol. 21, pp. 1069-1077, April 1, 2005 2005.
[24] A. T. Kwon, H. H. Hoos, and R. Ng, "Inference of transcriptional regulation relationships from gene expression data 10.1093/bioinformatics/btg106," Bioinformatics, vol. 19, pp. 905-912, May 22, 2003 2003.

[25] J. Qian, M. Dolled-Filhart, J. Lin, H. Yu, and M. Gerstein, "Beyond Synexpression Relationships: Local Clustering of Time-shifted and Inverted Gene Expression Profiles Identifies New, Biologically Relevant Interactions," J. Mol. Biol., pp. 1053-1066, 2001.
[26] H. D. Jong, "Modeling and Simulation of Genetic Regulatory Systems: A Literature Review," Journal of Computational Biology, vol. 9, pp. 67-103, 2002.
[27] T. Chen, "Modeling Gene Expression With Differential Equations," Pacific Symposium in Bioinformatics (PSB), World Scientific, vol. 4, pp. 29-40, 1999.
[28] S. Bulashevska and R. Eils, "Inferring genetic regulatory logic from expression data 10.1093/bioinformatics/bti388," Bioinformatics, p. bti388, March 22, 2005 2005.
[29] L. Mao and H. Resat, "Probabilistic representation of gene regulatory networks 10.1093/bioinformatics/bth236," Bioinformatics, vol. 20, pp. 2258-2269, September 22, 2004 2004.
[30] X.-w. Chen, G. Anantha, and X. Wang, "An effective structure learning method for constructing gene networks 10.1093/bioinformatics/btl090," Bioinformatics, vol. 22, pp. 1367-1374, June 1, 2006 2006.
[31] G. F. Cooper and E. Herskovits, "A Bayesian method for the induction of probabilistic networks from data," Machine Learning, vol. 9, pp. 309-347 1992.
[32] J. Suzuki, "A construction of Bayesian networks from databases based on an MDL scheme," Ninth Conference on Uncertainty in Artificial Intelligence, pp. 266-273, 1993.
[33] D. Husmeier, "Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks," Bioinformatics, vol. 19, pp. 2271-2282, 2003.
[34] P. Du, J. Gong, E. S. Wurtele, and J. A. Dickerson, "Modeling gene expression networks using fuzzy logic," IEEE Transactions on Systems, Man, and Cybernetics, vol. 35, pp. 1351-1359, 2005.
[35] J. Tuikkala, L. Elo, O. S. Nevalainen, and T. Aittokallio, "Improving missing value estimation in microarray data with gene ontology 10.1093/bioinformatics/btk019," Bioinformatics, p. btk019, December 23, 2005 2005.
[36] http://www.berkeley.edu/news/media/releases/2003/02/18_table.shtml
[37] http://www.union.wisc.edu/rna/newpics/bottom.jpg
[38] http://nanopedia.case.edu/image/dna.jpg
[39] http://alumni.media.mit.edu/ saul/PhD/imgs/ecoli.jpg
[40] http://www.med.monash.edu.au/assets/images/microbiology/microarray.jpg