Modeling Genetic Networks: Comparison of Static and Dynamic Models

Cristina Rubio-Escudero[1], Oscar Harari[1], Oscar Cordón[1,2], and Igor Zwir[1,3]

[1] Department of Computer Science and Artificial Intelligence,
[2] European Center for Soft Computing, Mieres, Spain.
[3] Howard Hughes Medical Institute, Washington University School of Medicine, St. Louis, MO.
{crubio, oharari, ocordon, zwir}@decsai.ugr.es

Abstract. Biomedical research has been revolutionized by high-throughput techniques and the enormous amount of biological data they are able to generate. The interest shown over network models and systems biology is rapidly raising. Genetic networks arise as an essential task to mine these data since they explain the function of genes in terms of how they influence other genes. Many modeling approaches have been proposed for building genetic networks up. However, it is not clear what the advantages and disadvantages of each model are. There are several ways to discriminate network building models, being one of the most important whether the data being mined presents a static or dynamic fashion. In this work we compare static and dynamic models over a problem related to the inflammation and the host response to injury. We show how both models provide complementary information and cross-validate the obtained results.

1 Introduction

Advances in molecular biology and computational techniques permit the systematical study of molecular processes that underlie biological systems (Durbin *et al*., 1998). One of the challenges of this post-genomic era is to know when, how and for how long a gene is turned *on/off*. Microarray technology has revolutionized modern biomedical research in this sense by its capacity to monitor the behavior of thousands of genes simultaneously (Brown *et al*., 1999; Tamames *et al.*, 2002). The reconstruction of genetic networks is becoming an essential task to understand data generated by microarray techniques (Gregory, 2005). The enormous amount of information generated by this high-throughput technique is raising the interest in network models to represent and understand biological systems.

Systems biology research arises at this point as the field to explore the life regulation processes in a cohesive way making use of the new technologies. Proteins have a main role in the regulation of genes (Rice and Stolovitzky, 2004), but unfortunately, for the vast majority or biological datasets available, there is no information about the level of protein activity. Therefore, we use the expression level of the genes as an indicator of the activity of proteins they generate.

Gene networks represent these gene interactions. A gene network can be described as a set of nodes which usually represent genes, proteins or other biochemical entities. Node interaction is represented with edges corresponding to biologic relations.

E. Marchiori, J.H. Moore, and J.C. Rajapakse (Eds.): EvoBIO 2007, LNCS 4447, pp. 78 – 89, 2007.

There is a wide range of models available to build genetic networks up. One of the differences between such models is whether they represent static or dynamic relations. Static modeling explains causal interactions by searching for mutual dependencies between the gene expression profiles of different genes (van Someren *et al.,* 2002). Clustering techniques are widely applied for static genetic network, since they group genes that exhibit similar expression levels.

In dynamic modeling, the expression of a node A in the network at time t_{+1} can be given as the result of the expression of the nodes in the network with edges related to A at time t (van Someren *et al.,* 2002). The understanding of the relations helps to describe all the relations occurring in a given organism we would be able to know the behavior of such organism throughout time.

The question arises as which network model is the most appropriate given a set of data. In the present work we have applied both static (*K*-means clustering method, (Duda and Hart, 1973)) and dynamic network models (a Boolean method, described in (D'onia *et al.*, 2003) and implemented in (Velarde, 2006) and a graphic Gaussian method (GGM) (Schäfer and Strimmer, 2005)) to a set of data derived from an experiment on inflammation and the host response to injury (Calvano *et al*., 2005). The results show how dynamic models are capable to recover temporal dependencies that static models are not able to find. Temporal studies are becoming widely used in biomedical research. In fact, over 30% of published expression data sets are time series (Simon *et al*., 2005).

2 Problem Description

In this work we compare the behavior of static vs. dynamic modeling in a problem derived from the inflammation and the host response to injury. On the one hand, static modeling searches for relations between the expression levels of genes throughout time. The relation found by static methods might not only be similar behavior throughout time (direct correlation), but an inverse correlation (two genes having exactly opposite profiles over time), a proximity on the expression values (distance measures such as Euclidean Distance or City block distance) (see Fig. 1). On the other hand, dynamic modeling retrieves temporal dependencies among genes, i.e., it detects dependencies of a gene at time t_{+1} related to some other(s) gene at time t (see Fig. 1).

To compare the performance of these two models, we have applied them to a data set derived from an experiment over inflammation and the host response to injury as part of a Large-scale Collaborative Research Project sponsored by the National Institute of General Medical Sciences (www.gluegrant.org) (Calvano *et al*., 2005). Human volunteers have been treated with intravenous endotoxin and compared to placebo, obtaining longitudinal blood expression profiles. Analysis of the set of gene expression profiles obtained from this experiment is complex, given the number of samples taken and variance due to treatment, time, and subject phenotype. The data were acquired from blood samples collected from eight human volunteers, four treated with intravenous endotoxin (i.e., patients 1 to 4) and four with placebo (i.e., patients 5 to 8). Complementary RNA was generated from circulating leukocytes at 0, 2, 4, 6, 9 and 24 hours after the and hybridized with GeneChips® HG-U133A v2.0 from Affymetrix Inc., which contains 22216 probe sets, analyzing the expression level of 18400 transcripts and variants, including 14500 well-characterized genes.

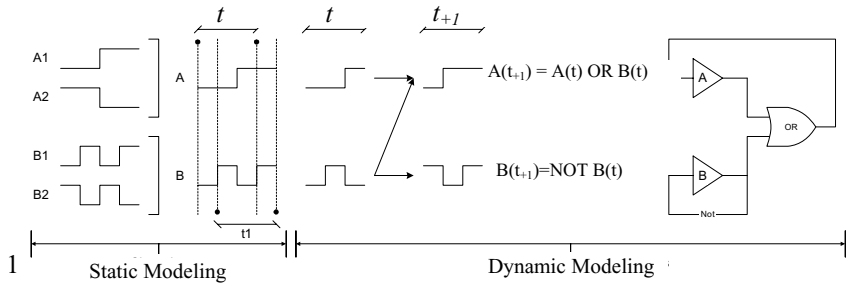

Fig. 1. The static modeling captures the relation (inverse correlation) between A_1 and A_2 (profile A) and between B_1 and B_1 (profile B). However, it does not capture the relation between A and B describing profile A at time t_{+1}. This relation is only captured by the dynamic model.

3 Genetic Network Construction

We have applied both static and dynamic models to the set of data just described. As said in Section 1, clustering techniques are widely applied for static genetic network, so we have used a classic clustering algorithm based on Euclidean distance, the *K-means* (Duda and Hart, 1973) which is a very popular clustering algorithm widely used with data from microarray experiments (Guiller *et al*., 2006). Two dynamic methods have been applied as well: a Boolean method, described in (D'onia *et al.*, 2003) and implemented in (Velarde, 2006) and a graphic Gaussian method (GGM) (Schäfer and Strimmer, 2005). These two methods have been chosen as representation of discrete and continuous models respectively, the two big families in which dynamic models can be divided (van Someren *et al.,* 2002). We now describe each of these methods.

Classification of gene expression patterns to explore shared functions and regulation can be accomplished using clustering methods (D'haeseleer *et al.*, 2000). We have applied a classic clustering algorithm based on Euclidean distance, the *K*-means algorithm (Duda and Hart, 1973). The number of resulting clusters *k* is estimated by application of the Davies-Bouldin validity index (Davies and Bouldin, 1979). The groupings obtained using this method, i.e., gene expression profiles, are expected to be functionally cohesive since genes sharing the same expression profiles are likely to be involved in the same regulatory process (D'haeseleer *et al.*, 2000). This can be proved applying the EMO-CC algorithm (Romero-Záliz *et al.,* 2006), which validates the gene groupings obtained using external information from the Gene Ontology database, which provides a controlled vocabulary to describe gene and gene product attributes in any organism (Ashburner *et al.,* 2000).

3.1 Dynamic Discrete Modeling : Boolean Networks

A Boolean network is composed by a set of nodes *n* which represent genes, proteins or other biochemical entities. These nodes can take *on/off* values. The net is determined by a set of at maximum *n* Boolean functions, each of them having the state of *k* specific nodes as input, where *k* depends on each node. Therefore, each node has its

own Boolean function which determines the next state based on the actual state of the input nodes. The changes in the net are assumed to occur at discrete time intervals.

The algorithm applied to build the Boolean network with our data is the GeneYapay (D'Onia *et al*., 2003). It performs an exhaustive search of Boolean functions over the data, where a number of nodes, less or equal then *k*, univocally determines the output of some other gene. All possible subsets of 1, 2, ..., *k* elements are visited calculating the number of inconsistencies of the Boolean functions in relation to the output value of each gene. The algorithm stops the search for each node when a subset of nodes is found which defines the expression profile. The implementation applied (Velarde, 2006) only uses the NAND function since all other Boolean function -AND, OR, NOT- can be expressed using NAND (see Table 1).

Table 1. Boolean functions obtained only using the NAND function

NOT A ≡ A NAND A
A AND B ≡ (A NAND B) NAND (A NAND B)
A OR B ≡ (A NAND A) NAND (B NAND B)

3.2 Dynamic Continuous Modeling : Graphic Gaussian Network

The graphical gaussian models were first proposed by Kishino and Waddell (2000) for the association structure among genes. GGMs are similar to Bayesian networks in that they allow to distinguish direct from indirect interactions (i.e. whether gene A acts on gene B directly or through a third gene C). As any graphical model, they also provide a notion of conditional independence of two genes. However, in contrast to Bayesian networks, GGMs contain only undirected rather than directed edges. This makes graphical Gaussian interaction modeling on the one hand conceptually simpler, and on the other hand more widely applicable (e.g. there are no problems with feedback loops as in Bayesian networks).

The GGM applied in this work has been developed by Schäfer and Strimmer, (2005) and is based on (1) improved (regularized) small-sample point estimates of partial correlation, (2) an exact test of edge inclusion with adaptive estimation of the degree of freedom and (3) a heuristic network search based on false discovery rate multiple testing.

4 Results

High-throughput techniques provide great amounts of data that need to be processed before being used to build genetic networks up. The first step is the identification of genes relevant for the problem under study. We have applied the methodology described in Rubio-Escudero *et al.,* (2005): a process based on the meta analysis of microarray data. The proliferation of related microarray studies by independent groups, and therefore, different methods, has lead to the natural step of combination of results (Gosh *et al.,* 2003). Thus, a battery of analysis methods has been applied (Student's T-Tests (Li and Wong, 2003), Permutation Tests (Tusher *et al*., 2001), Analysis of

Variance (Park *et al*., 2003) and Repeated Measures ANOVA (Der and Everitt, 2001)). A total of 2155 genes have been identified as relevant for the problem under study. For this particular problem the number of genes retrieved is very high compared to other microarray experiments, since the problem under study, inflammation and host response to injury, is a process that affects the human system in a global manner, hence altering the behavior of a large number of genes (Calvano *et al*., 2005).

At the view of these, we decide to use the expression profiles of the genes as the input for the genetic network building algorithms, since the number of genes involved in the problem is unfeasible for both building and analyzing the genetic networks. The set of profiles used is the one obtained from the static model applied, the K-*means* algorithm.

4.1 Static Modeling: *K*-Means Clustering

We apply a clustering method, the *K*-means algorithm, as described in section 3.1. We have identified 24 expression profiles (Rubio-Escudero *et al*., 2005) (see Fig. 2). These profiles have been proved as functionally cohesive by application of the EMO-CC algorithm (Romero-Záliz *et al.,* 2006). For instance, the majority of the genes exhibiting profile #22 are related to the inflammatory response (GO:0006954) and are annotated as intracellular (GO:0005622). Another sample is profile #16, with genes sharing the apoptosis (GO:0006915) and integral to plasma membrane (GO:0005887) annotations.

The functional identification of the 24 profiles resulting from the clustering method represents a further analysis of the data behind the identification of the genes relevant for the problem.

4.2 Dynamic Discrete Modeling: Boolean Network

Boolean building network algorithms use discrete data which take two possible values: *on* or *off*, i.e., 1 or 0. Therefore, the set of 24 differential profiles obtained in the inflammation and host response to injury problem (Calvano *et al*., 2005) needs to be transformed to fit the binary scheme. First of all, each of the profiles will be scaled in the [0, 1] interval according to the maximum value scored in the expression level of such profile throughout the six time points stored. The individual scaling has been used instead of a global one (scaling the 24 profiles according to the global maximum) since the profiles fluctuate in different levels of expression. For instance, profile #1 takes values between 1224.2 and 1724.4, while profile #24 changes between 13632 and 16436. If we scaled all values together, the variations between the expression values in profile #1 would result to small to be traceable, although they could be significative. In Table 2 (A) the expression levels before scaling are shown.

Once the values are scaled in the [0, 1] interval we have assign them [0-1] values. The simplest approach is to establish a threshold value, for instance 0.5, and to set each time point value depending whether they are over/under the threshold. The obvious problem with this approach is the “border value”, such as 0.45 or 0.55. These will be set 0 and 1 respectively, while they are so close to each other that they should take the same value. Our approach consists in setting the value based on the

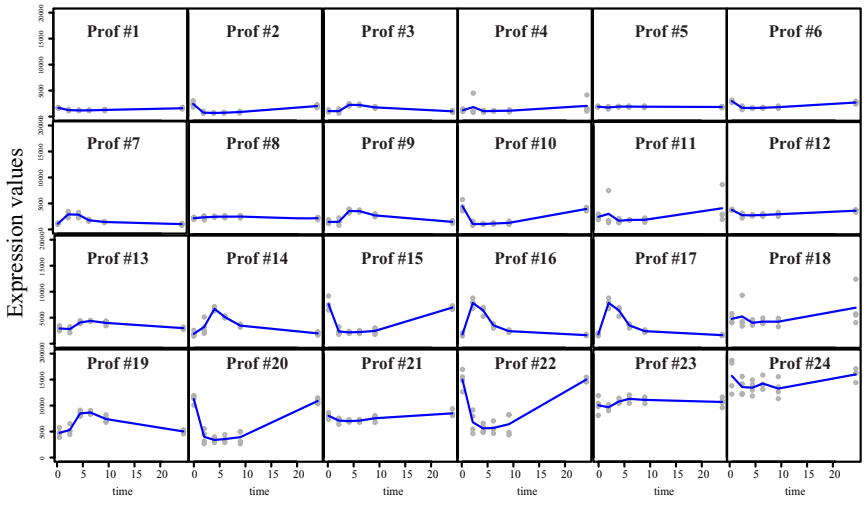

Fig. 2. Set of 24 expression profiles obtained from the inflammation and host response to injury problem

proximity to the expression level in the previous time point, which solves the previously described problem and captures the behavior of the profile over time. The scheme used to set the values is:

$$\begin{cases} if\ (|t - t_{+1}| < \delta) \quad then\ t_{+1} = t \\ if\ (|t - t_{+1}| > \delta) \quad then \begin{cases} if\ (|t - t_{+1}| < 0)\ then\ t_{+1} = 0 \\ if\ (|t - t_{+1}| > 0)\ then\ t_{+1} = 1 \end{cases} \end{cases}$$

where t_{+1} is the gene value to be set and t is the gene value in the previous time point. Table 2 (B) shows the obtained Boolean values for the 24 profiles in our problem.

The resulting Boolean network is shown in Fig. 3. This net is the result of an exhaustive search of Boolean functions over the data which univocally determines the output of the other genes. We see that some nodes represent more than one expression profile. This is due to the processing the data has to undergo. The scaling of the data to the [0, 1] interval, makes profiles at different levels of expression end up sharing a common Boolean profile. A sample of this in our particular problem is the one represented by profiles #9, #13 and #19. These three expression profiles share similar behaviour throughout time at different levels of expression (see Fig. 4). The net shows valuable information about relation between profiles. For instance, the relation established between profiles #7 and #17 with profiles #3 and #14 is confirmed when searching in the KEGG database (Kanehisa *et al.*, 2004), a metabolic pathway database. Genes exhibiting profiles #7 and #17 are in the same pathway and regulate

Table 2. Continuous and Boolean values obtained for each of the 24 profiles in the data set

PROFILES	CONTINUOUS VALUES (A)						BOOLEAN VALUES (B)					
	T0	T2	T4	T6	T9	T24	T0	T2	T4	T6	T9	T24
#1	1724.4	1316.4	1224.2	1236.9	1327.5	1666	1	0	0	0	0	1
#2	2546.2	734.44	700.28	737.5	867.44	2107.8	1	0	0	0	0	1
#3	1108.8	1027.9	2403.2	2376	1843.3	1069.6	0	0	1	1	0	0
#4	1323.6	2001.9	1089.4	1139.8	1192.7	2230.8	0	1	0	0	0	1
#5	1933.1	1829.8	1970.5	1983.6	1966.4	1907.5	1	0	1	1	1	1
#6	3146	1694.2	1669.1	1746.3	1889.8	2872.3	1	0	0	0	0	1
#7	1265.8	3551.7	3079	2008.1	1656.4	1160.3	0	1	1	0	0	0
#8	2396.3	2577.6	2721.5	2726.6	2712	2412.9	0	1	1	1	1	0
#9	1614.2	1619	3756.4	3972.6	3116.5	1676.8	0	0	1	1	1	0
#10	4844.2	1278.3	1248.4	1316.9	1468.1	4240.1	1	0	0	0	0	1
#11	2730.3	3351.4	1921.3	2114.9	2146.3	4459.3	0	1	0	0	0	1
#12	4176	2984.1	2974	3068.7	3265.5	4021.8	1	0	0	0	0	1
#13	3022.8	2898.1	4262.2	4666.1	4329.1	3150.8	0	0	1	1	1	0
#14	2117.6	3289.7	7298.8	5871.3	4036.8	2229.4	0	0	1	1	0	0
#15	7849.5	2328	2297.4	2450	2738.6	7171.7	1	0	0	0	0	1
#16	4836.6	4220.5	5085.4	5398.3	5356.3	4829.7	1	0	1	1	1	0
#17	1950.7	9001.6	7946	4268.8	2804	1787.1	0	1	1	0	0	0
#18	5238.2	5734.5	4445.8	4654.6	4665.7	7584.4	1	0	0	0	0	1
#19	4935.7	5335.4	9034.5	9171	7858	5285.3	0	0	1	1	1	0
#20	11615	4161.2	3578.6	3760.8	4149.9	11344	1	0	0	0	0	1
#21	8358.3	7308.8	7244.2	7652.2	8139.2	8913.8	1	0	0	0	0	1
#22	15442	7021.5	5798.9	5918.8	6632.3	15605	1	0	0	0	0	1
#23	10473	10132	11396	11871	11531	10980	0	0	1	1	1	1
#24	16095	13749	13632	14364	13741	16436	1	0	0	1	0	1

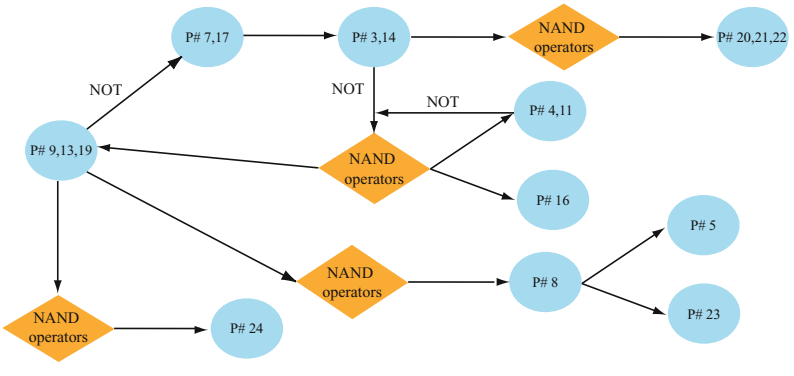

Fig. 3. Genetic network obtained using the Boolean model. The round nodes represent the gene expression profiles (groups of genes with a common behavior) and the diamond shape nodes represent the Boolean function based on the NAND operator. Note that some nodes represent more than one expression profile.

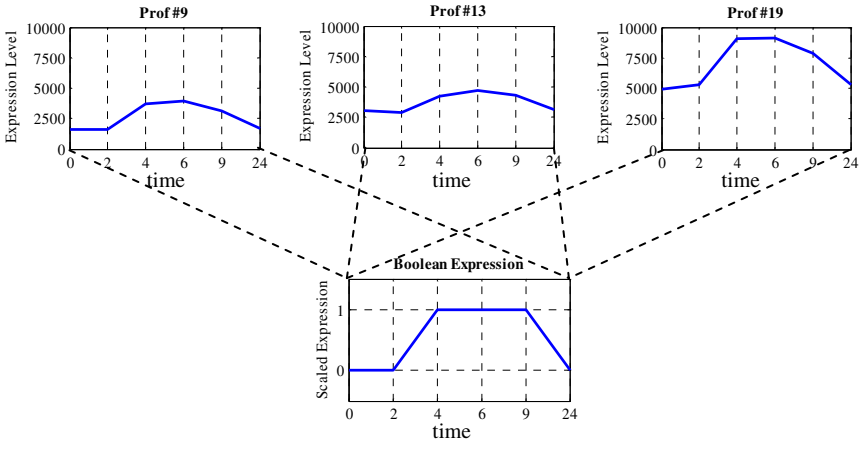

Fig. 4. Profiles at different levels of expression but sharing a common behavior throughout time share the same Boolean profile

genes exhibiting profile #14 (See Fig. 5). That is the case of gene IL1RN (prof. #17, Interleukin-1 receptor antagonist protein precursor), related to the immune response (GO:0006955) and gene IL1R2 (prof. #14, Interleukin-1 receptor type II), also related to the immune response. We can see in Fig. 5(A) more examples of gene relations found in KEGG and present in the Boolean network obtained.

4.3 Dynamic Continuous Modeling: Graphical Gaussian Model

We have applied a Graphic Gaussian algorithm (Schäfer and Strimmer, 2005), which takes as input continuous data that can be in longitudinal format (Opgen-Rhein and Strimmer, 2006), very convenient for microarray time course experiments since it deals with repeated measurements, irregular sampling, and unequal temporal spacing of the time points. To select the edges, and thus the nodes, we have used the local false discovery rate (fdr) (expected proportion of false positives among the proposed edges), an empirical Bayes estimator of the false discovery rate (Efron, 2005). An edge is considered *present* or *significant* if its local fdr is smaller than 0.2 (Efron, 2005). Three independent networks are found (see Fig. 6). Network (B) confirms the information provided by the Boolean network about profiles #7, #14 and #17. In network (A) there is a relation established between profiles #11, #23 and #16 that is confirmed when searching in the KEGG database (see Fig. 5(B)). That is the case of gene RACK (Reversion-inducing cysteine-rich protein with Kazal motifs), which exhibits profile #11 and is related to gene MMP9 (Matrix metalloproteinase-9), which exhibits profile #23. Both genes are related to the inflammation problem. Another relation is found between a gene exhibiting profile #23, CEBPB (CCAAT/enhancer-binding protein beta), related to the immune response (GO:0006955) and to the; inflammatory response (GO:0006954) and a gene exhibiting profile #16, CASP1 (Caspase-1) related to apoptosis (GO:0006915).

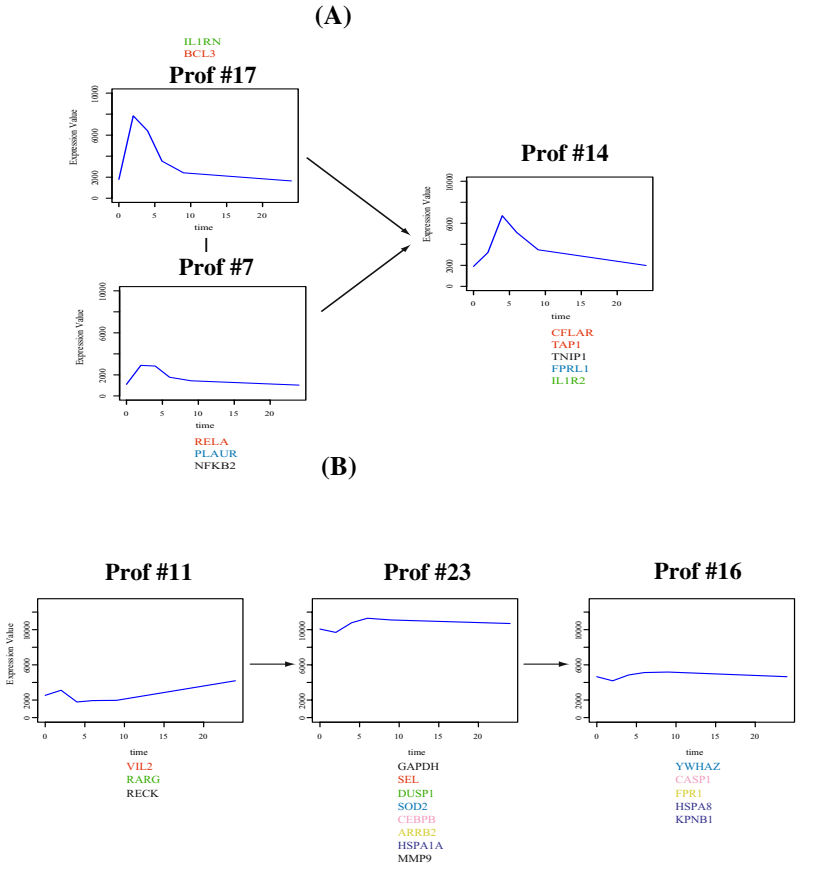

Fig. 5. Gene relations detected by the network building algorithms and confirmed in the KEGG database. (A) has been found by both the Boolean algorithms and GGM while (B) has only been found by GGM. The genes regulate other genes with the same color.

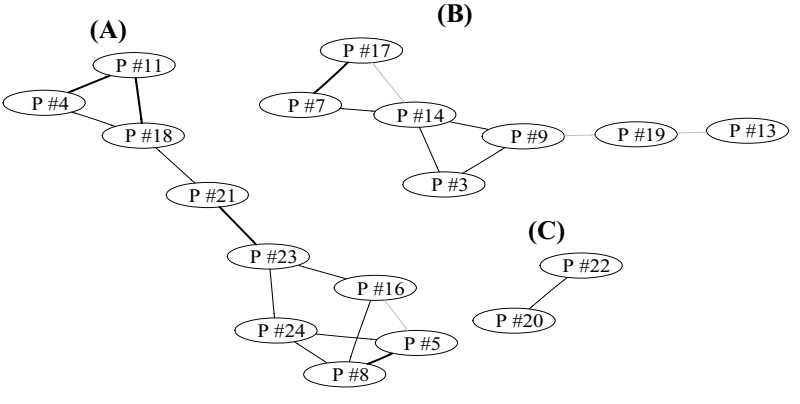

Fig. 6. Three independent networks found by the GGM algorithm

5 Discussion

We have applied both static and dynamic methods for the analysis of a data set derived from the inflammation and the host response to injury (Calvano *et al*., 2005). The static method has been the *K*-means clustering algorithm, and the dynamic methods have been a discrete one, Boolean model described in (D'Onia *et al*., 2003) and implemented by (Velarde, 2006) , and a continuous one, Graphic Gaussian Model developed by (Schäfer and Strimmer, 2005). We have already described some of the findings these methods have made on the dataset: the static method is capable of grouping the genes based on their behaviour throughout time and these groupings are cohesive in biological functionality. The dynamic models provide temporal relations between the genes, or in this case, between the profiles they exhibit, organizing them in regulatory networks that are validated using the KEGG database. These temporal relations would not have been found only applying static models.

When comparing the two dynamic models, we see that they cross-validate in general their results i.e., the profiles involved and the relations between those profiles are concordant with one another. The Boolean algorithm and GGM show different and complementary information about the problem under study. In a GGM network the relation between nodes is based on the levels of correlation but the time dependency is not so clearly pointed out as in Boolean networks. For instance, in our GGM net we see that profiles #5, #8 and #23 are related since they are in the same subnet, but the Boolean network specifically describes the behavior of those profiles: #8 determines the behavior of both #5 and #23 (see Fig. 7), since the behavior shown by profile #8 is shifted over time in profiles #5 and #23. This kind of information is only available in network models which strongly stress the temporal dependencies, as it is the case with Boolean networks.

However, Boolean algorithms lack the capacity to distinguish among expression profiles with similar behaviour throughout time at different levels of expression (see Fig. 4). For instance, the Boolean algorithm considers profiles #9, #13 and #19 as only one node. GGM uses continuous values solving this problem and taking advantage of the diversity or the data, but it misses some information. The network (C) provided by GGM covers profiles #20 and #22. In the Boolean network they are considered as one single profile along with #21, since their Boolean representation is the same. GGM has not been able to capture the similarity between these three profiles, only between two of them, #20 and #22. However, the Boolean model considers them as the same node, so any temporal relation between them is impossible to capture. In fact, when searching in KEGG (Kanehisa *et al.*, 2004), we see that one of the genes that exhibit profile #20 is NFKB2 (nuclear factor of kappa light polypeptide gene enhancer in B-cells 2) and one of the genes exhibiting profile #22 is TNIP1 (TNFAIP3-interacting protein 1). When searching for information about these two genes, which are related in their behavior, we see they are also functionally related since TNIP1 interacts with zinc finger protein A20/TNFAIP3 and inhibits TNF-induced NF-kappa-B-dependent gene expression (NFKB2). This valuable information is only prone to be found with network models such as GGM which permit the representation of temporal dependencies among strongly correlated profiles.

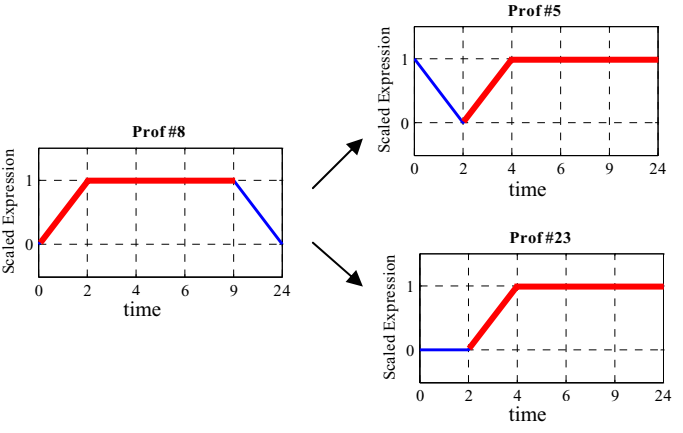

Fig. 7. Time relations found by the Boolean algorithm. Profile #8 determines the behavior of profiles #6 and #23.

The evaluation of static and dynamic models over the inflarmmation and host response to injury problem allows us to conclude that static models provide very valuable information but a step further is needed to get a deeper knowledge of the problem under study. Dynamic models provide information of the temporal dependencies in the data what is very valuable especially for time-course experiments, which are becoming very popular used in biomedical research. Dynamic discrete models miss valuable information when discretizing the data, while the continuous models do not suffer this problem. However, dynamic continuous models are not capable to find some of the dependencies that discrete model discover and vice versa. Therefore, they are complementary methods and it is a recommendable practice to apply both models to extract the maximum information possible from experiments.

References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, M., Rubin, G.M. and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat Genet*, 25: 25-29.

Brown,P. and Botstein,D. (1999) Exploring the new world of the genome with DNA microarrays. *Nature Genet.*, 21 (Suppl.), 33-37.

Calvano,S.E., Xiao,W., Richards,D.R., Feliciano,R.M., Baker, H.V., Cho, R.J., Chen, R.O., Brownstein,B.H., Cobb,J.P., Tschoeke,S.K., Miller-Graziano,C., Moldawer,L.L., Mindrinos, M.N., Davis, R.W., Tompkins,R.G. and Lowry,S.F. (2005) The Inflammation and Host Response to Injury Large Scale Collaborative Research Program. A Network-Based Analy-sis of Systemic Inflammation in Humans. *Nature*, 13:437(7061):1032-7.

Davies, D. L. and Bouldin, W. (1979) A cluster separation measure. IEEE PAMI, 1, 224-227.

D'haeseleer P., Liang S. and Somogyi R. (2000) Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics*, 16(8):707-726.

D'Onia D., Tam L., Cobb J. P., and Zwir I. A hierarchical reverse-forward methodology for learning complex genetic networks. *Proceedings of the 3rd International Conference on Systems Biology (ICSB)*, Stockholm Sweden.

Duda, R. O., and Hart, P. E. (1973) Pattern Classification and Scene Analysis. John Wiley & Sons, New York, USA.

Durbin,R., Eddy,S., Krogh,A. and Mitchison,G. (1998) Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press.

Efron, B. (2005). Local false discovery rates. Preprint, Dept. of Statistics, Stanford University.

Dutilh B. (1999) Analysis of data from microarray experiments, the state of the art in gene network reconstruction. *Report Binf.1999.11.01, Bioinformatics*, Utrecht University.

Gregory W. (2005) Inferring network interactions within a cell. *Bioinformatics*, 6: 380-389.

Guiller A., Bellido A., Coutelle A. and Madec L. (2006) Spatial genetic pattern in the land Helix aspersa inferred from a 'centre-based clustering' procedure. Genet Res., 88(1):27-44.

Kanehisa M., Goto S., Kawashima S., Okuno Y. and Hattori M. (2004) The KEGG resource for deciphering the genome. Nucleic Acids Res, 1;32(Database issue):D277-80.

Karp P.D., Ouzounis C.A., Moore-Kochlacs C., Goldovsky L., Kaipa P., Ahren D., Tsoka S., Darzentas N., Kunin V. and Lopez-Bigas N. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Research, 19:6083-89 2005.

Kishino,H. and Waddell,P.J. (2000) Correspondence analysis of genes and tissue types and finding genetic links from microarray data. Genome Informatics, 11:83–95.

McAdams H. H. and Arkin A. (1998) Simulation Of Prokaryotic Genetic Circuits. Ann. Rev. Biophys. Biom. Struct., 27:199-224.

Opgen-Rhein R.and Strimmer K. (2006) Inferring gene dependency networks from genomic longitudinal data: a functional data approach. REVSTAT 4:53-65.

Rice J.J. and Stolovitzky G. (2004) Making the most of it: Pathway reconstruction and integrative simulation using the data at hand. Biosilico 2(2):70-7.

Romero-Záliz R., Rubio-Escudero C., Cordón O., Harare O., del Val C. and Zwir I. (2006) Mining Structural Databases: An Evolutionary Multi-Objective Conceptual Clustering Methodology. Proceedings of the 4th European Workshop on Evolutionary Computation and Machine Learning in Bioinformatics. Budapest, Hungary.

Rubio-Escudero C., Romero-Záliz R., Cordón O., Harari O., del Val C., Zwir I. (2005) Optimal Selection of Microarray Analysis Methods using a Conceptual Clustering Algorithm. Proceedings of the 4th European Workshop on Evolutionary Computation and Machine Learning in Bioinformatics. Budapest, Hungary.

Schäfer J. and Strimmer K. (2005) An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics 21: 754-764.

Silvescu A. and Honavar V. (2001) Temporal Boolean Network Models of Genetic Networks and Their Inference from Gene Expression Time Series. Complex Systems, 13(1):54-75.

Simon I., Siegfried Z., Ernst J. and Bar Z. (2005) Combined static and dynamic analysis for determining the quality of time-series expression profiles. Nat. Biotechnol., 23(12):1503-8.

Velarde, Cyntia. (2006) Master Thesis in Comp Sci, University of Buenos Aires, Argentina.