Learning and Validating Bayesian Network Models of Genetic Regulatory Networks

Jose M. Peña
Department of Physics and Measurement Technology
Linköping University, S-581 83 Linköping, Sweden
jmp@ifm.liu.se

Abstract

We propose a framework for learning from data and validating Bayesian network models of genetic regulatory networks. The learning phase selects multiple locally optimal models of the data and reports the best of them. The validation phase assesses the confidence in the model reported by studying all the different locally optimal models obtained in the learning phase. We prove that our proposal is asymptotically optimal under the faithfulness assumption. Experiments with real data (320 samples of the expression levels of 32 genes involved in *Saccharomyces cerevisiae*, i.e. baker's yeast, pheromone response) show that our proposal is reliable.

1 Introduction

Much of a cell's complex behavior can be explained through the concerted activity of genes and gene products. This concerted activity is typically represented as a network of interacting genes and gene products. Identifying this network, which we call *genetic regulatory network* (GRN), is crucial for understanding the behavior of the cell which, in turn, can lead to better diagnosis and treatment of diseases. This is one of the most exciting challenges in computational biology. For the last few years, there has been an increasing interest in using *Bayesian networks* (BNs) to model GRNs (Friedman et al., 2000; Hartemink et al., 2002; Ott et al., 2004; Pe'er et al., 2001), mainly due to their ability to represent stochastic relations between genes. This is particularly important when inferring models of GRNs from gene expression data, because gene expression seems to be a stochastic phenomenon (Blake et al., 2003; McAdams and Arkin, 1997), and because gene expression data typically include measurement noise. Another key feature of BNs is the existence of principled ways for learning them from gene expression data, and even for incorporating prior knowledge in the learning process.

A BN models a GRN by representing a joint probability distribution for V, $p(V)$, where V denotes a nonempty finite set of discrete random variables such that each of them represents the expression level of a gene in the GRN. Formally, a BN for V is a pair (G, θ), where G is an acyclic directed graph (DAG) whose nodes correspond to the random variables in V, and θ are parameters specifying a conditional probability distribution for each node $X \in V$ given its parents $Pa(X)$ in G, $p(X|Pa(X))$. A BN represents $p(V)$ through the factorization $p(V) = \prod_{X \in V} p(X|Pa(X))$.

Once a BN is built as the model of a GRN, it constitutes an effective device for reasoning. In particular, the DAG G of the BN allows to reason about the structure of $p(V)$ in terms of conditional (in)dependencies among the random variables in V. These conditional (in)dependencies can be read from G by means of the d-separation criterion (Lauritzen, 1996). In this paper, we focus on this sort of structural reasoning. We define a BN *model*, $M(G)$, as the set of all the joint probability distributions that satisfy the conditional independencies enforced by the d-separation criterion in G. In other words, $M(G)$ contains all the joint pro-

bability distributions that can be represented by all the BNs with structure G, i.e. all the possible parameterizations of G. Joint probability distributions that do not satisfy any other conditional independence than those enforced by the d-separation criterion in G are called *faithful* to G. We are concerned here with BN models of GRNs. Furthermore, we aim to induce these models from gene expression data automatically. Two main approaches to learning BN models from data exist: One tests conditional independencies among the random variables in V with the help of the learning data (Spirtes et al., 1993), while the other searches the space of models by scoring them with respect to the learning data (Castelo and Kočka, 2003; Chickering, 2002; Nielsen et al., 2003). In this paper, we take the latter approach, also called *model selection*.

Model selection aims to find the highest scoring BN model of the learning data. Unfortunately, this task is NP-complete (Chickering, 1996).[1] For this reason, most model selection algorithms are heuristic and they only guarantee convergence to a locally optimal model. As the number of locally optimal models can be huge (Nielsen et al., 2003), validating the model learnt is crucial. When inferring BN models of GRNs from gene expression data, validation becomes even more important: Gene expression databases usually contain few samples and these may even be very noisy. However, there is little research on BN model validation.

In this paper, we propose a framework for learning from data and validating BN models of GRNs. The learning phase consists in running repeatedly a stochastic model selection algorithm to discover multiple locally optimal models of the learning data and, then, reporting the best of them. The validation phase assesses the confidence in some features of the model reported by studying all the different locally optimal models obtained in the learning phase. The higher the confidence in the features of the model reported, the more reliable or valid it is. We prove that our framework is asymptotically optimal under the faithfulness assumption.[2]

The paper is structured as follows. We describe the learning and validation phases in Sections 2 and 3, respectively. We evaluate our proposal with real data in Section 4. The data consist of 320 samples of the expression levels of 32 genes involved in *Saccharomyces cerevisiae* (baker's yeast) pheromone response. We conclude in Section 5 with some discussion.

2 Learning Phase

As mentioned above, the learning phase runs repeatedly a stochastic model selection algorithm to obtain multiple locally optimal models of the learning data and, then, reports the best of them. We use the *k-greedy equivalence search algorithm* (KES) (Nielsen et al., 2003) for this purpose. Like most model selection algorithms, KES consists of three components: A *neighborhood*, a *scoring criterion* and a *search strategy*. The neighborhood of a model restricts the search to a small part of the search space around the model, and it is usually defined by means of local transformations of a representative of the model. The scoring criterion evaluates the quality of a model with respect to the learning data. The search strategy selects a new model, based on the scoring criterion, from those in the neighborhood of the current best model. The paragraphs below describe these components in the case of KES.

KES uses the *inclusion boundary* of a model as the neighborhood of the model. The inclusion boundary of a model M_1, $IB(M_1)$, is the union of the *lower* and *upper* inclusion boundaries, $LIB(M_1)$ and $UIB(M_1)$, respectively. $LIB(M_1)$ is the set of models M_2 that are strictly included in M_1 and such that no model strictly included in M_1 strictly includes M_2. Likewise, $UIB(M_1)$ is the set of models M_2 that strictly include M_1 and such that no model strictly including M_1 is strictly included in M_2. $IB(M_1)$ is characterized using DAGs

[1] To be exact, the result reported in (Chickering, 1996) states that identifying the BN model that maximizes the Bayesian scoring criterion when computed as indicated in (Heckerman et al., 1995) is NP-complete. It is usually assumed that this result holds for other common scoring criteria as well, though there is not yet a formal proof.

[2] By asymptotically optimal we mean optimal in the large sample limit.

as the set of models represented by all those DAGs that can be obtained by adding or removing a single edge from any representative DAG of M_1 (Chickering, 2002). Any representative DAG G_1 of a model M_1 can be obtained from any other representative DAG G_2 of M_1 through a sequence of *covered* edge reversals in G_2, where the edge $X \rightarrow Y$ is covered in G_2 if $Pa(Y) = Pa(X) \cup \{X\}$ (Chickering, 2002).[3]

In this paper, KES uses the Bayesian information criterion (Schwarz, 1978) as the scoring criterion. The Bayesian scoring criterion can be considered as well. KES uses the following search strategy:

```
KES (k∈[0,1])
M = empty graph model
repeat
   B = set of models in IB(M) with
       higher score than the model M
   if |B| > 0 then
      C = random subset of the set B
          with size max(1,|B|·k)
      M = the highest scoring model
          from the set C
   else return(M)
```

where $|B|$ denotes the cardinality of the set B. The input parameter $k \in [0, 1]$ allows to trade off greediness for randomness. This makes KES ($k \neq 1$) able to reach different locally optimal models when run repeatedly. KES ($k = 1$) corresponds to the greedy equivalence search algorithm (GES) proposed in (Chickering, 2002).[4] We refer the reader to (Nielsen et al., 2003) for a thorough description and study of KES, including the proof of the following property.

Theorem 1 *KES using fully observed learning data i.i.d. sampled from a joint probability distribution faithful to a DAG G asymptotically always returns $M(G)$.*

[3] A more efficient, though more complex, characterization of $IB(M_1)$ using completed acyclic partially directed graphs is reported in (Studený, 2003a; Studený, 2003b).

[4] To be exact, GES is a two-phase algorithm that first uses only $UIB(M)$ and, then, only $LIB(M)$. KES ($k = 1$) corresponds to a variant of GES described in (Chickering, 2002) that uses $IB(M)$ in each step.

3 Validation Phase

As noted in the introduction, the best locally optimal model discovered in the learning phase may not represent perfectly the distribution of the learning data, because these may be noisy, sparse and/or very complex. This is usually the case when dealing with gene expression data. Validating the best model found in the learning phase is, then, of much importance. Our proposal for validating it consists of two main steps. First, extraction of relevant features from the model. Second, assessment of the confidence in the features extracted. The higher the confidence in these features, the more believable or valid the model is. The following sections describe the two steps of the validation phase.

3.1 Feature Extraction

First of all, we need to adopt a model representation scheme so that interesting features can be extracted and studied. Representing a model by a DAG may not the best solution, because there may be many such representative DAGs of the model. A *completed acyclic partially directed graph* (CPDAG) provides, on the other hand, a canonical representation of a model. A CPDAG represents a model by summarizing all its representative DAGs: The CPDAG contains the directed edge $X \rightarrow Y$ if and only if $X \rightarrow Y$ exists in all the representative DAGs, while it contains the undirected edge X—Y if and only if $X \rightarrow Y$ exists in some representative DAGs and $Y \rightarrow X$ in some others. In this paper, we use CPDAGs as the model representation scheme. See (Chickering, 2002) for an efficient procedure to transform a DAG into its corresponding CPDAG.

We pay attention to four types of features in a CPDAG: Directed edges, undirected edges, directed paths and *Markov blanket neighbors* (two nodes are Markov blanket neighbors if there is an edge between them, or if they are both parents of another node). We focus on these types of features because they stress relevant aspects of the distribution of the learning data. Directed and undirected edges reflect unmediated interactions between random variables. In addi-

tion, directed edges suggest possible causal relations. Directed paths establish orderings between random variables. A random variable is conditionally independent of all the random variables outside its Markov blanket neighborhood given its Markov blanket neighborhood (Lauritzen, 1996). Moreover, the Markov blanket neighborhood of a random variable is the minimal set with such a property.

3.2 Confidence Assessment

While all the different locally optimal models discovered in the learning phase disagree in some features, we expect them to share some others. In fact, the more strongly the learning data supports a feature, the more frequently it should appear in the different locally optimal models found. Likewise, the more strongly the learning data supports a feature, the higher the likelihood of the feature being true in the distribution that generated the learning data. This leads us to assess the confidence in a feature as the fraction of models containing the feature out of all the different locally optimal models obtained in the learning phase. Note that we give equal weight to all the models available, no matter their scores. Alternatively, we could weight each of the models by its score. These two approaches to confidence estimation are close in spirit to the methods proposed in (Friedman et al., 2000; Hartemink et al., 2002; Pe'er et al., 2001). No proof of optimality is reported for these methods. However, we prove below that our proposals are asymptotically optimal under the faithfulness assumption.

Theorem 2 *Assessing the confidence in a feature as the (weighted) fraction of models containing the feature out of the different locally optimal models obtained by running KES repeatedly using fully observed learning data i.i.d. sampled from a joint probability distribution faithful to a DAG G asymptotically always assigns confidence equal to one to the features in $M(G)$ and equal to zero to the rest.*

Proof: Under the conditions of the theorem, KES asymptotically always returns $M(G)$ (Theorem 1). □

3.3 Validity Assessment

Let M^* denote the best model found in the learning phase. Deciding on the validity of M^* based only on the confidence values scored by its features may be difficult. We suggest a sensible way to ease making this decision. We call *true positives* (TPs) to the features in M^* with confidence value equal or above a given threshold value t. Likewise, we call *false positives* (FPs) to the features not in M^* with confidence value equal or above t, and *false negatives* (FNs) to the features in M^* with confidence value below t. In order to decide on the validity of M^*, we propose studying the trade-off between the number of FPs and FNs for each type of features under study as a function of t. The less FPs and FNs for high values of t, the more believable or valid M^* is. In other words, we trust M^* as a valid model of the learning data if the features in M^* receive high confidence values, while the features not in M^* score low confidence values. Note that we treat on equal basis FPs and FNs. Alternatively, we can attach different costs to FPs and FNs according to our preferences (e.g. we may be less willing to accept FPs than FNs).

Finally, it is worth noting that, under the faithfulness assumption, M^* asymptotically always coincides with the true model (Theorem 1) and the number of FPs and FNs is asymptotically always zero for any threshold value $t > 0$ (Theorem 2). Therefore, under the faithfulness assumption, the learning phase asymptotically always returns the true model and the validation phase asymptotically always confirms the validity of this model.

4 Evaluation

In this section, we evaluate our methodology for learning from data and validating BN models of GRNs. We first introduce the gene expression database used for the evaluation and the experimental setting. Then, we report and discuss the results obtained.

4.1 Database and Setting

We use the database in (Hartemink et al., 2002) for the evaluation. The database consists of 320

Group	Description
Magenta	Genes expressed only in MATa cells: STE2, MFA1, MFA2, STE6, AGA2 and BAR1
Red	Genes expressed only in MATα cells: STE3, MFALPHA1, MFALPHA2 and SAG1
Blue	Genes whose promoters are bound by Ste12: FUS3, STE12, FAR1, FUS1 and AGA1
Green	Genes coding for components of the heterotrimeric G-protein complex: GPA1, STE4 and STE18
Yellow	Genes coding for core components of the signaling cascade (except FUS3 which is blue): STE7, STE11 and STE5
Orange	Genes coding for auxiliary components of the signaling cascade: KSS1, STE20 and STE50
Brown	Genes coding for components of the SWI-SNF complex: SNF2 and SWI1
White	Others: SST2, KAR3, TEC1, MCM1, SIN3 and TUP1

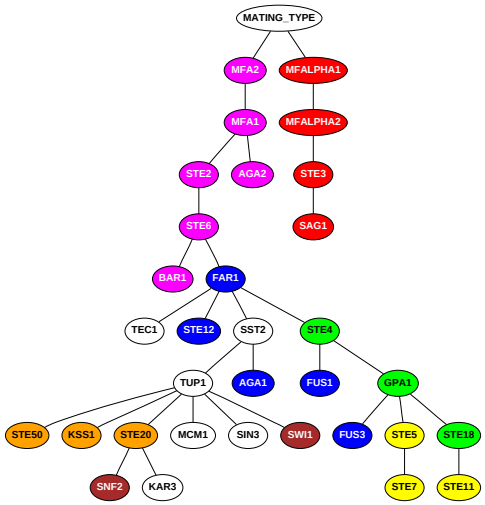

Figure 1: Left, description of the groups of genes in the evaluation. Right, best model learnt represented as a CPDAG. Nodes are colored with the color of the group they belong to.

records with each record being characterized by 33 attributes. The records correspond to 320 samples of unsynchronized *Saccharomyces cerevisiae* (baker's yeast) populations observed under different experimental conditions. Yeast is considered an ideal eukaryotic organism and, thus, it has been widely studied (Lee et al., 2002; Spellman et al., 1998). The first 32 attributes of each record represent the expression levels of 32 genes involved in yeast pheromone response. This pathway plays an essential role in the sexual reproduction of yeast. The last attribute of each record, named MATING_TYPE, indicates the mating type of the strain of yeast in the corresponding sample, either MATa or MATα, as some of the 32 genes measured express only in strains of a specific mating type. We note that gene expression levels are discretized into four states. We refer the reader to (Hartemink et al., 2002) for details on the data collection and preparation process, as well as for a thorough description of the 32 genes in the database. We summarize this description in Figure 1 (left) by grouping the genes according to their function in the domain under study.

The setting for the evaluation is as follows. We consider KES ($k = 0.6, 0.8, 0.9$). We avoid values of k too close to 0 so as to prevent convergence to poor locally optimal models (Nielsen et al., 2003). For each value of k, we proceed in two phases, learning and validation, as previously described in Sections 2 and 3, respectively. We first run KES 1000 independent times and report the best locally optimal model found. Then, we use all the different locally optimal models discovered in the 1000 runs to estimate the confidence in the features of interest, i.e. directed and undirected edges, directed paths and Markov blanket neighbors. Finally, we compute the trade-off between the number of FPs and FNs for each type of features under study as a function of the threshold value t, in order to decide on the validity of the model reported. We give equal weight to all the models used for confidence estimation, because most of them have similar scores. We treat equally FPs and FNs when computing the trade-off. We use the current knowledge of the domain under study to check the consistency and accuracy of the results obtained.

4.2 Results

We first report the results of the learning phase. Out of the 1000 independent runs of KES performed for each value of k considered in the evaluation, we obtained 967 different locally optimal models for $k = 0.6$, 330 for $k = 0.8$, and 159 for $k = 0.9$. In the three cases, the best

t	$k = 0.6$		$k = 0.8$		$k = 0.9$	
	FPs	FNs	FPs	FNs	FPs	FNs
1.00	0	30	0	25	0	22
0.95	0	22	0	15	0	12
0.90	0	17	0	11	0	10
0.85	0	12	0	8	0	7
0.80	0	11	0	6	0	3
0.75	0	8	0	2	0	1
0.70	0	5	0	1	0	1
0.65	0	2	0	1	0	1
0.60	0	1	0	0	0	0
0.55	0	1	0	0	0	0
0.50	0	0	0	0	0	0
0.45	0	0	0	0	0	0
0.40	0	0	0	0	0	0
0.35	0	0	0	0	0	0
0.30	1	0	0	0	0	0
0.25	6	0	0	0	0	0
0.20	9	0	4	0	2	0
0.15	11	0	7	0	6	0
0.10	17	0	11	0	10	0
0.05	25	0	18	0	14	0

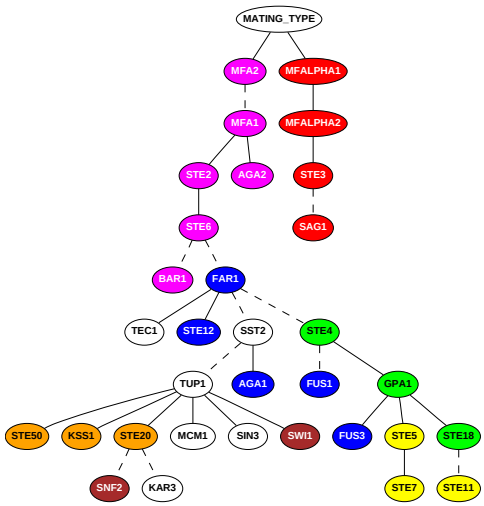

Figure 2: Left, trade-off between the number of FPs and FNs for undirected edges at threshold values $t = 0.05 \cdot r$, $r = 1, \ldots, 20$. Right, undirected edges for $k = 0.8$ when $t = 0.60$ (solid and dashed edges) and when $t = 0.90$ (solid edges). Nodes are colored with the color of the group they belong to.

model found was the same. Figure 1 (right) shows the best model learnt depicted as a CPDAG. We remark that the graph in the figure does not intend to represent the biological or physical GRN (e.g., gene products are not modelled), but the conditional (in)dependencies in it. We note that all the edges in the CPDAG in the figure are undirected, meaning that each edge appears in opposite directions in at least two representative DAGs of the model. As a matter of fact, none of CPDAGs representing the locally optimal models obtained in the 3000 runs of KES performed has directed edges. This reduces the types of features under study to one: Undirected edges.

We now discuss the results of the validation phase. Figure 2 (left) shows the trade-off between the number of FPs and FNs for undirected edges as a function of the threshold value t. We note that the CPDAG corresponding to the best model found in the learning phase has 32 undirected edges. As can be appreciated from the figure for each value of k considered in the evaluation, FNs only happen for high values of t, while FPs only occur for low values of t. Therefore, TPs receive substantially higher confidence values than FPs. For $k = 0.8$, for instance, no TP scores lower than 0.60, while no FP scores higher than 0.25. These observations support the validity and meaningfulness of the best model discovered in the learning phase. Figure 2 (right) depicts the undirected edges for $k = 0.8$ when $t = 0.60, 0.90$. Note that all the edges in the figure are TPs. As a matter of fact, there are 0 FPs and 0 FNs (32 TPs) for $t = 0.60$, and 0 FPs and 11 FNs (21 TPs) for $t = 0.90$. The figures for $k = 0.6, 0.9$ are very similar to the one shown. We omit them due to space restrictions.

It is worth mentioning that we repeated the experiments in this section with a random database created by randomly reshuffling the entries of each attribute in the original database. In such a database, we did not expect to find features scoring high confidence values. As a matter of fact, no edge was added in any of the 3000 runs of KES performed. This leads us to believe that the results presented above are not artifacts of the learning and validation phases but reliable findings.

We give below some evidence that the conditional (in)dependencies in the best model induced in the learning phase are consistent with the existing knowledge of yeast pheromone response. This somehow confirms the results of

the validation phase. Magenta-colored genes are marginally dependent one on another. Moreover, no genes from other groups mediate in these dependencies. These observations also hold true for red- and green-colored genes. These findings are consistent with the genes in each of these groups being functionally related, as indicated in Figure 1 (left). Moreover, magenta- and red-colored genes are also marginally dependent on MATING_TYPE, and this is the only node that mediates between both groups. This makes sense given that these two groups of genes express in strains of yeast of different mating type. Orange-colored genes are marginally dependent one on another, and only TUP1 mediates in these dependencies. As a matter of fact, TUP1 has the highest number of adjacencies in the model, which is consistent with its role as repressor of numerous genes in yeast pheromone response (Hartemink et al., 2002). Blue-colored genes are marginally dependent one on another. However, several other genes mediate in these dependencies. These observations also hold true for yellow- and brown-colored genes. Many of the mediating genes have also been identified playing such a role in (Hartemink et al., 2002). The following adjacencies are also supported by the existing literature: STE2—STE6, STE3—SAG1, SST2—AGA1, MFALPHA2—STE3, MFA1—AGA2, FAR1—TEC1 and STE6—FAR1 (Spellman et al., 1998), and TUP1—MCM1 (Gavin et al., 2000). We defer a more thorough discussion on the biological validation to an extended version of this paper.

Finally, it is worth mentioning that most of the undirected edges scoring high confidence values in the validation phase are supported by the existing knowledge of yeast pheromone response. For instance, most undirected edges in Figure 2 (right) with confidence values equal of above 0.90 have been discussed in the paragraphs above. Therefore, we can conclude that the framework proposed in this paper for learning from data and validating BN models of GRNs is accurate and reliable: The learning phase has produced a model that is consistent with the existing knowledge of the domain under study, and the validation phase has confirmed, independently of the existing knowledge, that the model is indeed meaningful.

5 Discussion

Learning BN models of GRNs from gene expression data is a challenging task. On one hand, learning BN models from data is difficult in itself (NP-complete and highly multimodal). On the other hand, gene expression databases are noisy and sparse. For these reasons, many of the works in the literature focus on extracting features (directed and undirected edges, directed paths and Markov blanket neighbors) from the learning data rather than on model selection (Friedman et al., 2000; Hartemink et al., 2002; Pe'er et al., 2001). In order to ease visualization, the features extracted for each of the types under study are usually arranged in a graph, with an edge between two nodes if they are related by a feature with high confidence. In our opinion, a major limitation of this approach is that this graph does not represent a (global) model of the distribution of the learning data, but a collection of (local) patterns because each feature corresponds to a piece of local information. As a consequence, this graph does not, in general, allow reasoning about the conditional (in)dependencies in the distribution of the learning data, which is a major drawback. For instance, the lack of an edge between two nodes in the graph does not necessarily mean that they are marginally or conditionally independent. Moreover, the types of features typically considered are very local, as they only relate pairs of nodes but not triplets or sets of nodes, and deciding when a feature scores sufficient confidence is rather arbitrary. In this paper, we approach the problem of learning BN models of GRNs from gene expression data in a rather different way: We suggest selecting a single model of the learning data and, then, validating it. The validation step aims to assess whether the model selected is reliable or not as a whole. When the model fails to be reliable, our proposal reduces to the one discussed above, and the model is interpreted as a set of patterns.

The experimental results reported in this paper show that our framework for model learning and validation is accurate and consistent.

Another major difference between this paper and the works cited above is the neighborhood used within the model selection algorithm. While they consider classical neighborhoods based on local transformations (e.g. single edge addition, removal or reversal) of a single representative DAG of the current best model, we use the inclusion boundary neighborhood which takes into account every single representative DAG of the current best model in order to generate the set of neighboring models. The inclusion boundary neighborhood outperforms the classical neighborhoods in practice without compromising the runtime, because it reduces the risk of getting stuck in a locally but not globally optimal model (Castelo and Kočka, 2003). Furthermore, the inclusion boundary neighborhood allows to derive important theoretical results about asymptotic optimal learning of BN models from data (Castelo and Kočka, 2003; Chickering, 2002; Nielsen et al., 2003). These results guarantee that the framework for learning from data and validating BN models of GRNs that we present in this paper is asymptotically optimal under the faithfulness assumption. This result does not hold when using any of the classical neighborhoods within the model selection algorithm.

A line of research we are engaged in consists in using the results of the validation phase to design gene perturbations, gather new data and refine the models obtained in the learning phase accordingly. Combining observational and interventional data will also provide insight into the causal relations in the GRN under study.

Acknowledgements

We thank Alexander J. Hartemink for providing the data used in this work.

References

Blake, W. J., Kærn, M., Cantor, C. R. and Collins, J. J.: Noise in Eukaryotic Gene Expression. Nature **422** (2003) 633-637.

Castelo, R. and Kočka, T.: On Inclusion-Driven Learning of Bayesian Networks. Journal of Machine Learning Research **4** (2003) 527-574.

Chickering, D. M.: Learning Bayesian Networks is NP-Complete. In Learning from Data: Artificial Intelligence and Statistics V (1996) 121-130.

Chickering, D. M.: Optimal Structure Identification with Greedy Search. Journal of Machine Learning Research **3** (2002) 507-554.

Friedman, N., Linial, M., Nachman, I. and Pe'er, D.: Using Bayesian Networks to Analyze Expression Data. Journal of Computational Biology **7** (2000) 601-620.

Gavin, I. M., Kladde, M. P. and Simpson, R. T.: Tup1p Represses Mcm1p Transcriptional Activation and Chromatin Remodeling of an a-Cell-Specific Gene. The EMBO Journal **19** (2000) 5875-5883.

Hartemink, A. J., Gifford, D. K., Jaakkola, T. S. and Young, R. A.: Combining Location and Expression Data for Principled Discovery of Genetic Regulatory Network Models. In Pacific Symposium on Biocomputing **7** (2002) 437-449.

Heckerman, D., Geiger, D. and Chickering, D. M.: Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. Machine Learning **20** (1995) 197-243.

Lauritzen, S. L.: Graphical Models. Clarendon Press, Oxford (1996).

Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph, Z., Gerber, G. K., Hannett, N. M., Harbison, C. T., Thompson, C. M., Simon, I., Zeitlinger, J., Jennings, E. G., Murray, H. L., Gordon, D. B., Ren, B., Wyrick, J. J., Tagne, J. B., Volkert, T. L., Fraenkel, E., Gifford, D. K. and Young, R. A.: Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. Science **298** (2002) 799-804.

McAdams, H. H. and Arkin, A.: Stochastic Mechanisms in Gene Expression. In Proceedings of the National Academy of Science of the USA **94** (1997) 814-819.

Nielsen, J. D., Kočka, T. and Peña, J. M.: On Local Optima in Learning Bayesian Networks. In Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence (2003) 435-442.

Ott, S., Imoto, S. and Miyano, S.: Finding Optimal Models for Small Gene Networks. In Pacific Symposium on Biocomputing **9** (2004) 557-567.

Pe'er, D., Regev, A., Elidan, G. and Friedman, N.: Inferring Subnetworks from Perturbed Expression Profiles. Bioinformatics **17** (2001) 215-224.

Schwarz, G.: Estimating the Dimension of a Model. Annals of Statistics **6** (1978) 461-464.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. and Futcher, B.: Comprehensive Identification of Cell Cycle-Regulated Genes of the Yeast *Saccharomyces cerevisiae* by Microarray Hybridization. Molecular Biology of the Cell **9** (1998) 3273-3297.

Spirtes, P., Glymour, C. and Scheines, R.: Causation, Prediction, and Search. Springer-Verlag, New York (1993).

Studený, M.: Characterization of Inclusion Neighbourhood in Terms of the Essential Graph: Upper Neighbours. In Proceedings of the Seventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (2003a) 161-172.

Studený, M.: Characterization of Inclusion Neighbourhood in Terms of the Essential Graph: Lower Neighbours. In Proceedings of the Sixth Workshop on Uncertainty Processing (2003b) 243-262.