Information-Theoretic Inference of Gene Networks Using Backward Elimination

Patrick E. Meyer[1,2,3], **Daniel Marbach**[1,2], **Sushmita Roy**[1,2] **and Manolis Kellis**[1,2]
[1]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, MA, USA
[2]Broad Institute of MIT and Harvard, MA, USA
[3]Machine Learning Group, Université Libre de Bruxelles, Brussels, Belgium

Abstract— *Unraveling transcriptional regulatory networks is essential for understanding and predicting cellular responses in different developmental and environmental contexts. Information-theoretic methods of network inference have been shown to produce high-quality reconstructions because of their ability to infer both linear and non-linear dependencies between regulators and targets. In this paper, we introduce MRNETB an improved version of the previous information-theoretic algorithm, MRNET, which has competitive performance with state-of-the-art algorithms.*

MRNET infers a network by using a forward selection strategy to identify a maximally-independent set of neighbors for every variable. However, a known limitation of algorithms based on forward selection is that the quality of the selected subset strongly depends on the first variable selected. In this paper, we present MRNETB, an improved version of MRNET that overcomes this limitation by using a backward selection strategy followed by a sequential replacement. Our new variable selection procedure can be implemented with the same computational cost as the forward selection strategy.

MRNETB was benchmarked against MRNET and two other information-theoretic algorithms, CLR and ARACNE. Our benchmark comprised 15 datasets generated from two regulatory network simulators, 10 of which are from the DREAM4 challenge, which was recently used to compare over 30 network inference methods. To assess stability of our results, each method was implemented with two estimators of mutual information. Our results show that MRNETB has significantly better performance than MRNET, irrespective of the mutual information estimation method. MRNETB also performs comparably to CLR and significantly better than ARACNE indicating that our new variable selection strategy can successfully infer high-quality networks.

Keywords: mutual information, systems biology, network inference

1. Introduction

Transcriptional regulatory networks are networks of transcription factor (TF) proteins and their target genes, where each edge represents the regulatory activity of each TF on a target gene. These networks summarize the underlying regulatory circuitry by which precise, condition-specific patterns of gene expression are generated under different cellular contexts. Understanding these networks can not only provide insights into how cells function, but can lead to effective drug delivery for treating human diseases by specific network-informed intervention.

Network inference methods (also called *reverse engineering* methods) attempt to reconstruct the transcriptional network of a cell from gene expression data [1]. However, this is a challenging task because of the large amount of experimental and biological noise in expression data, and because of the high dimensionality and combinatorial nature of the problem [25]. Notwithstanding, the bioinformatics community has developed several successful network inference techniques in the last decade [8]. In particular, information-theoretic methods for network inference have been shown to scale to datasets with several thousand genes and a limited amount of samples [7], [13], [15].

In this paper, we present MRNET Backward (MRNETB), an improved version of the previous network inference method called MRNET (Minimum Redundancy NETwork) [15]. MRNET infers a network of interactions between TFs and target genes by using a forward selection strategy to identify a maximally independent set of neighbors for every variable. However, methods based on forward selection, including MRNET, rely on the correct choice of the first neighbor and suffer in performance if the first neighbor is chosen incorrectly. MRNETB overcomes this limitation by implementing a combination of a backward elimination and a sequential replacement procedure. Importantly, we implement our new neighbor selection procedure at the same computational cost as forward selection.

Our proposed technique is benchmarked against MRNET and two other state-of-the-art information-theoretic algorithms, namely CLR (Context Likelihood of Relatedness) [7] and ARACNE (Algorithm for the Reconstruction of Accurate Cellular Networks) [13]. Our benchmark comprised fifteen datasets generated from two regulatory network simulators. Ten out of the fifteen datasets come from the DREAM4 challenge, a competition devoted to compare the performance of network inference methods [11]. To assess stability of our results, each method was implemented

with two different estimators of mutual information. Our results show that irrespective of the mutual information estimation used, MRNETB performs better than MRNET and ARACNE and performs comparably against CLR. Our new strategy thus successfully infers high-quality networks, without incurring additional computational costs.

The rest of the paper is organized as follows: Section 2 provides background on information-theoretic methods of network inference and the two network inference methods, CLR and ARACNE. Section 3 describes MRNET and our new approach, MRNETB. Section 4 describes the experimental framework and results and the conclusions are summarized in Section 5.

2. Mutual Information Network Inference

Mutual information network inference methods comprise a subcategory of network inference methods, which infer regulatory interactions between genes based on pairwise mutual information [15]. As a first step, these methods require the computation of the mutual information matrix (MIM), a square matrix whose mim_{ij} element is given by the mutual information between X_i and X_j:

$$\text{mim}_{ij} = I(X_i; X_j) \tag{1}$$

where X_i and X_j are random variables denoting the expression levels of genes i and j, respectively.

The main advantages of mutual information networks for inference of transcriptional regulatory networks are:

- The computational complexity is affordable. This results from the fact that only $\binom{n}{2}$ calls of mutual information, based on bivariate probability distributions, are required to compute the MIM [13].
- The number of required samples is rather low, since only bivariate distribution are to be estimated [17].

In the following, we first review two state-of-the-art network inference methods based on pairwise mutual information. We proceed by describing two commonly used mutual information estimators.

2.1 Mutual Information Estimation

Two popular ways of computing mutual information are: (1) discretizing variables with an equal frequency binning (so that marginal distributions are uniform), and (2) assuming normally distributed variables. We will now describe these two procedures.

2.1.1 Uniformly distributed discrete variables and empirical estimation

If X_i is a continuous random variable taking values between a and b, the interval $[a, b]$ can be discretized by partitioning it into $|\mathcal{X}_i|$ subintervals, called *bins*, where the symbol $\mathcal{X}_i$ denotes the bin vector. We use $\#(x_{i_k})$ to denote the number of data points in the kth bin of variable X_i, the symbol $m = \sum_k \#(x_{i_k})$ to denote the total number of samples and the symbol $\hat{p}(x_{i_k}) = \frac{\#(x_{i_k})}{m}$ to denote the empirical probability.

One possible binning strategy is to divide the interval $[a, b]$ of X_i such that each subinterval contains the same number of data points. It follows that subinterval sizes are typically different. This binning strategy is often referred as *equal frequency binning*. The number of subintervals should be chosen so that all bins contain a significant number of samples. Here, we set $|\mathcal{X}_i| = \sqrt{m}$, which is a common choice [28]. The entropy of X_i is $H(X_i) = \log |\mathcal{X}_i|$, since its distribution is uniform across the bins. As a result, the "empirical" or "plug-in" estimation of the mutual information becomes

$$\begin{array}{c} I(X_i; X_j) = \log(|\mathcal{X}_i||\mathcal{X}_j|) \\ + \sum_{x_{i_k} \in \mathcal{X}_i} \sum_{x_{j_l} \in \mathcal{X}_j} \hat{p}(x_{i_k}, x_{j_l}) \log \hat{p}(x_{i_k}, x_{j_l}) \end{array} \tag{2}$$

It should be noted that this entropy estimation tends to underestimate the true value of entropy, i.e. $E[H(p)] \leq H(E[p])$ [18].

2.1.2 Normally distributed continuous variables

Let X be a multivariate Gaussian random variable with mean μ and covariance C. The probability density function of X is

$$f(X) = \frac{1}{\sqrt{(2\pi)^n |C|}} \exp^{(-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu))}. \tag{3}$$

where $|C|$ is the determinant of the covariance matrix.

The mutual information between two Gaussian random variables, X_i and X_j, is then given by

$$I(X_i, X_j) = \frac{1}{2} \log \left(\frac{\sigma_{ii}\sigma_{jj}}{|C|} \right) \tag{4}$$

$$= -\frac{1}{2} \log(1 - \rho^2). \tag{5}$$

where ρ is the Pearson's correlation [9].

The Spearman rank correlation was shown to have better empirical performance for network inference than Pearson's correlation [17].

2.2 Context Likelihood of Relatedness (CLR)

The CLR algorithm [7] is an extension of the relevance network approach [4]. The latter approach has been introduced for gene clustering and successfully applied to infer relationships between RNA expression and chemotherapeutic susceptibility [3]. The relevance networks approach consists of inferring a network in which a pair of genes $\{X_i, X_j\}$ are linked by an edge if the mutual information $I(X_i; X_j)$ is larger than a given threshold θ. The complexity of the method is $O(n^2)$ since all pairwise interactions are considered.

The CLR algorithm derives a score from the empirical distribution of the mutual information for each pair of genes. In particular, instead of considering the information $I(X_i; X_j)$ between genes X_i and X_j, it estimates a score $w_{ij} = \sqrt{z_i^2 + z_j^2}$, where

$$z_i = \max\left(0, \frac{I(X_i; X_j) - \mu_i}{\sigma_i}\right) \quad (6)$$

The parameters μ_i and σ_i are the mean and the standard deviation of the empirical distribution of the mutual information values $I(X_i, X_k)$ of X_i with all other variables X_k ($k = 1, \ldots, n$). The CLR algorithm has a complexity in $O(n^2)$. It was successfully applied to decipher the *E. coli* transcriptional regulatory network [7].

2.3 Algorithm for Reconstruction of Accurate Cellular Networks (ARACNE)

ARACNE [13] is based on the Data Processing Inequality, which states that, if gene X_i interacts with gene X_j through gene X_k, then

$$I(X_i; X_j) \leq \min\left(I(X_i; X_k), I(X_j; X_k)\right)$$

ARACNE first assigns to each pair of nodes a weight equal to their mutual information. Then, all edges for which $I(X_i; X_j) < \theta$ are removed, where θ is a given threshold. Eventually, the weakest edge of each triplet is interpreted as an indirect interaction and is removed.

An extension of ARACNE removes the weakest edge only if the difference between the two lowest weights lies above a threshold η. Hence, η allows the number of pruned edges to be tunable.

If the network is a tree including only pairwise interactions, the method guarantees the reconstruction of the original network, once it is provided with the exact MIM [13]. ARACNE's complexity is $O(n^3)$ since the algorithm considers all triplets of genes. ARACNE successfully recovered components of transcriptional regulatory networks in mammalian cells and was shown to have favorable performance compared to Bayesian networks and relevance networks [13].

3. Minimum Redundancy Networks

MRNET infers a network by using a variable selection procedure called Maximum Relevance Minimum Redundancy (MRMR) for every random variable $X_j \in X$, [23], [19]. Assume X_j is the variable for which we need to select the predictor variables. The MRMR methods ranks a set $X_{S_j} \subseteq \{X \setminus X_j\}$ of the predictor variables according to the difference between the mutual information of $X_i \in X_{S_j}$ with X_j (the relevance) and the average mutual information with the selected variables in X_{S_j} (the redundancy). The rationale is that direct interactions should be ranked before indirect interactions (i.e. the ones with redundant information with the direct ones) by the method.

The MRMR method has been introduced together with a forward selection that starts by selecting the variable X_k that has the highest mutual information with the target X_j. The second selected variable X_i will be the one having a high information $I(X_i; X_j)$ with the target and at the same time a low information $I(X_i; X_k)$ with the previously selected variable. In the following steps, given a set X_{S_j} of selected variables, X_{S_j} is updated by choosing the variable, X_i^{MRMR}, that maximizes the MRMR score s_i:

$$\begin{gathered} X_i^{MRMR} = \arg\max_{X_i \in X_{-(i,j)}} (s_i) \\ s_i = I(X_i; X_j) - \frac{1}{|S_j|} \sum_{k \in S_j} I(X_i; X_k) \end{gathered} \quad (7)$$

At each step of the algorithm, the selected variable thus represents a trade-off between relevance and redundancy.

The generic principle of MRNET consists of identifying a subset X_{S_j} (for each variable X_j, $j = 1, 2.., n$) whose variables have maximal pairwise relevance with X_j and, at the same time, maximal pairwise independence among them. More formally,

$$\begin{gathered} X_{S_j} = \arg\max_{X_{S_j} \in X_{-j}} (u - r) \\ u = \frac{1}{|S_j|} \sum_{i \in S_j} I(X_i; X_j) \\ r = \frac{2}{|S_j|(|S_j|-1)} \sum_{i,k>i \in S_j} I(X_i; X_k) \end{gathered} \quad (8)$$

The network inference approach MRNET consists of repeating this selection procedure for each target gene $X_j \in X$. For each pair $\{X_i, X_j\}$, MRNET returns two (not necessarily equal) scores s_i and s_j according to (7). The score of the pair $\{X_i, X_j\}$ is defined as the maximum between s_i and s_j. A specific network can then be inferred by only keeping edges whose score lies above a given threshold θ (as in CLR). Thus, the algorithm infers an edge between X_i and X_j either when X_i is a well-ranked predictor of X_j ($s_i > \theta$), or when X_j is a well-ranked predictor of X_i ($s_j > \theta$). In practice, the selection of variables stops when the average redundancy term $\frac{1}{|S_j|} \sum_{k \in S_j} I(X_i; X_k)$ exceeds the relevance term $I(X_i; X_j)$.

MRNET has complexity of $O(n^3)$ since each variable selection is $O(n^2)$. MRNET was shown to have similar or better performance than alternative information-theoretic network inference methods [15], [10], [21], [17].

3.1 MRNET Backward (MRNETB)

In this section, we introduce MRNETB, an improved version of MRNET. This new algorithm overcomes a major limitation of MRNET coming from the use of forward selection as the subset search strategy. A known limitation of algorithms based on forward selection, including MRNET, is that the quality of the selected subset strongly depends on the first variable selected. If the first selected variable, i.e. the variable with the highest mutual information with the target variable X_j, is not a true neighbor of X_j, then minimizing

redundancy (or equivalently maximizing independency) with that wrongly selected variable is not desirable.

The optimisation problem of MRNET (8) is a binary quadratic optimization problem. Backward elimination combined with a sequential search is known to perform well on binary quadratic problems [2]. The backward elimination method starts with a set containing all the variables and then selects the variable X_i whose removal induces the highest increase of the objective function till the stopping criterion is fulfilled (i.e. a relevance term $\frac{1}{|S_j|}\sum_{i\in S_j} I(X_i;X_j)$ higher than redundancy term $\frac{2}{|S_j|(|S_j|-1)}\sum_{i,k>i\in S_j} I(X_i;X_k)$). The procedure is enhanced by an iterative sequential replacement which, at each step, swaps the status of a selected and a non selected variable such that the largest increase in the objective function is achieved. The sequential replacement is stopped when no further improvement is met.

Forward selection, backward elimination, and sequential replacement all have an algorithmic complexity of $O(n^2)$ [14]. In other words, the network built using a backward elimination followed by sequential replacement has the same asymptotic computational cost as the one based on a forward selection strategy. Backward elimination has been previously shown to outperform forward selection in variable selection [6]. In the next section, we show that the backward selection strategy also improves the performance in network inference using mutual information.

We have made MRNETB available in the latest version (2.5) of our bioconductor open-source package called minet [16].

4. Experiments

We compare the performance of our new approach MRNETB with the three existing network inference methods described above (ARACNE, CLR and MRNET) using a framework based on artificial (simulated) regulatory networks (Fig 1). In simulation, the ground truth is known and inferred networks can be systematically evaluated (which is typically not possible in vivo) [11].

The framework is composed of the following four steps:

1) Produce artificial gene expression data from networks of known structure
2) Compute the MIM using two different mutual information estimators for each produced dataset, namely the empirical and the Spearman based estimator.
3) Infer the network from each computed MIM
4) Assess the quality of the inferred networks using the area under the precision-recall (PR) curve.

4.1 Metrics

The area under the PR curve (AUPRC) is a standard metric used in the network inference community [20]. The PR curve plots precision (pre) against recall (rec) for different values of a threshold determining which pairs qualify as a predicted edge [22]. Precision, pre, is given by

$$pre = \frac{tp}{tp + fp} \tag{9}$$

Precision measures the fraction of real edges (tp) among the ones classified as positive ($tp + fp$). Recall, rec, also called true positive rate (tpr), is given by

$$rec = tpr = \frac{tp}{tp + fn} \tag{10}$$

It denotes the fraction of real edges that are correctly inferred.

In a network inference setting, the PR curve illustrates the trade-off between eliminating many low confidence edges by using a high threshold (low recall, high precision) versus keeping many arcs (high recall) with doubt on their significance (low precision). These quantities depend on the threshold chosen. A good network inference method maximizes both precision and recall.

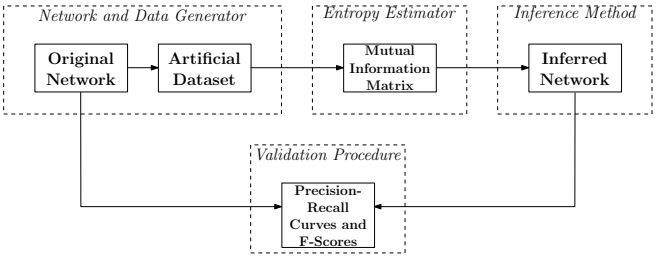

Fig. 1: Experimental framework used to assess network reconstruction quality

4.2 Datasets

We compared the different network inference methods on fifteen datasets (Table 1). These datasets come from two different generators, namely Syntren [27] and GeneNetWeaver (GNW) [12], [11]. Ten out of these fifteen datasets are from the DREAM4 challenge, which was recently used to compare over 30 network inference methods. Briefly, five of the DREAM4 datasets are from the so-called multifactorial challenge. These five datasets have been obtained by applying multifactorial perturbations to the original network. The other five datasets contain time courses capturing the response of the network to a perturbation, followed by gradual return to steady state upon removal of the perturbation. For a detailed description of these datasets, refer to the DREAM project website (http://wiki.c2b2.columbia.edu/dream/).

4.3 Results

We compared MRNETB against three state-of-the-art information theoretic methods for network inference (Tables 2 and 3) using the area under the PR curve (AUPRC) on the benchmark datasets, as described in the previous section. MRNETB has significantly higher AUPRC than both

Dataset	Source	Variables	Samples
syntren1	Syntren - E.Coli	300	100
syntren2	Syntren - E.Coli	300	100
syntren3	Syntren - E.Coli	300	100
syntren4	Syntren - E.Coli	300	100
syntren5	Syntren - E.Coli	300	100
multifact1	Dream4-Multifact1	100	100
multifact2	Dream4-Multifact2	100	100
multifact3	Dream4-Multifact3	100	100
multifact4	Dream4-Multifact4	100	100
multifact5	Dream4-Multifact5	100	100
dream1	Dream4-TS1	100	210
dream2	Dream4-TS2	100	210
dream3	Dream4-TS3	100	210
dream4	Dream4-TS4	100	210
dream5	Dream4-TS5	100	210

Table 1: Benchmark Datasets

MRNET and ARACNE (p-values were computed using a paired Wilcoxon test on the AUPRC, see Table 4), suggesting that MRNETB outperforms these methods. MRNETB outperforms CLR on the DREAM time series datasets, and is outperformed by CLR on the Syntren datasets, suggesting that MRNETB and CLR have at par performance. Note that the results for ARACNE are over-pessimistic because the method typically produces networks with high precision and low recall. As a result, for lower thresholds the area under ARACNE's PR-curve is much smaller than for the other methods.

The results are similar for both estimators of mutual information (i.e. empirical and Spearman based estimation). However, network inference based on Spearman's correlation has higher performance than inference based on empirical estimation of mutual information.

It is worth noting that MRNETB and CLR would have ranked third and fourth in the multifactorial DREAM4 challenge.

	CLR	ARACNE	MRNET	MRNETB
syntren1	**0.12**	0.04	0.09	0.11
syntren2	**0.12**	0.02	0.08	0.10
syntren3	**0.12**	0.02	0.08	0.10
syntren4	**0.11**	0.02	0.07	0.09
syntren5	**0.12**	0.02	0.08	0.10
multifact1	0.12	0.12	0.11	**0.13**
multifact2	0.12	0.11	0.11	**0.13**
multifact3	**0.23**	0.19	0.21	**0.23**
multifact4	**0.18**	0.14	0.14	**0.18**
multifact5	0.16	0.15	0.16	**0.17**
dream1	**0.11**	0.05	0.08	0.10
dream2	**0.11**	0.08	0.10	**0.11**
dream3	**0.18**	0.10	0.12	**0.18**
dream4	**0.14**	0.08	0.10	0.13
dream5	0.13	0.09	0.12	**0.14**
avg	0.138	0.082	0.11	0.133

Table 2: AUPRC Scores for the Empirical Estimation of Mutual Information. Best scores are in bold.

	CLR	ARACNE	MRNET	MRNETB
syntren1	**0.13**	0.04	0.10	0.12
syntren2	**0.13**	0.03	0.10	0.12
syntren3	**0.14**	0.04	0.11	0.12
syntren4	**0.13**	0.04	0.10	0.12
syntren5	**0.14**	0.03	0.09	0.12
multifact1	**0.22**	0.18	0.21	0.21
multifact2	0.21	0.16	0.22	**0.24**
multifact3	0.34	0.35	**0.38**	0.32
multifact4	0.33	0.29	**0.34**	0.31
multifact5	0.32	0.32	**0.36**	0.30
dream1	**0.09**	0.05	0.08	**0.09**
dream2	0.10	0.08	0.11	**0.12**
dream3	0.18	0.11	0.18	**0.21**
dream4	**0.12**	0.07	0.11	**0.12**
dream5	0.14	0.11	**0.17**	0.15
avg	0.181	0.126	0.177	0.178

Table 3: AUPRC Scores for the Squared Spearman Rho Correlation. Best scores are in bold.

MRNETB Losses/Ties/Wins	CLR	ARACNE	MRNET
Total	16/6/8	2/0/28	4/0/26
P-val MRNETB vs	CLR	ARACNE	MRNET
Total	0.11	5e-6	0.01

Table 4: Losses/ties/wins of MRNETB vs each method and p-values obtained with a paired Wilcoxon test.

5. Conclusion

This paper introduces MRNETB, a new information-theoretic method for inferring gene regulatory networks using gene expression data. Similar to other state-of-the-art information-theoretic methods, MRNETB relies on pairwise mutual information. As a result, this method can tackle datasets with large number of variables and low number of samples. An appealing aspect of the new method is the use of a better and more robust search for identifying a maximally informative subset of variables than a classic forward selection, without incurring additional computational cost. We compared MRNETB against three other approaches on fifteen datasets and the experimental results show that the proposed technique is competitive. Further research will focus on using MRNETB within an integrative framework to assess the relative information contribution of different data sources for improving network reconstruction.

References

[1] A. Ambesi-Impiombato and D. di Bernardo. Computational biology and drug discovery: from single-target to network drugs. *Current Bioinformatics*, 2006.

[2] A. Billionnet and F. Calmels. Linear programming for the 0-1 quadratic knapsack problem. *European Journal of Operational Research*, 92:310–325, 1996.

[3] A. J. Butte, P. Tamayo, D. Slonim, T.R. Golub, and I.S. Kohane. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proceedings of the National Academy of Sciences*, 97(22):12182–12186, 2000.

[4] A. J. Butte and I. S. Kohane. Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurments. *Pacific Symposium on Biocomputing*, 5:415–426, 2000.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley, New York, 1990.

[6] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research* **3** (2003) 1157–1182

[7] J.J. Faith, B. Hayete, J.T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J.J. Collins, and T.S. Gardner. Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biology*, 5, 2007.

[8] T. S. Gardner and J. Faith. Reverse-engineering transcription control networks. *Physics of Life Reviews 2*, 2005.

[9] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall International, 1999.

[10] F. M. Lopes, D. C. Martins, and R. M. Cesar. Comparative study of grns inference methods based on feature selection by mutual information. In *IEEE International Workshop on Genomic Signal Processing and Statistics*, 2009.

[11] D. Marbach, R. J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, and G. Stolovitzky. Revealing strengths and weaknesses of methods for gene network inference. *PNAS*, 107(14):6286-6291, 2010.

[12] D. Marbach, T. Schaffter, C. Mattiussi, and D. Floreano. Generating realistic in silico gene networks for performance assessment of reverse engineering methods. *Journal of Computational Biology*, 16(2):229-239, 2009.

[13] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. Dalla Favera, and A. Califano. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7, 2006.

[14] P. Merz and B. Freisleben. Greedy and local search heuristics for unconstrained binary quadratic programming. *Journal of Heuristics*, 8(2):1381–1231, 2002.

[15] P. E. Meyer, K. Kontos, F. Lafitte, and G. Bontempi. Information-theoretic inference of large transcriptional regulatory networks. *EURASIP Journal on Bioinformatics and Systems Biology*, Special Issue on Information-Theoretic Methods for Bioinformatics, 2007.

[16] P. E. Meyer, F. Lafitte, and G. Bontempi. Minet: An open source r/bioconductor package for mutual information based network inference. *BMC Bioinformatics*, 2008.

[17] C. Olsen, P. E. Meyer, and G. Bontempi. On the impact of entropy estimation on transcriptional regulatory network inference based on mutual information. *EURASIP Journal on Bioinformatics and Systems Biology*, 2009.

[18] L. Paninski. Estimation of entropy and mutual information. *Neural Computation*, 15(6):1191–1253, 2003.

[19] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

[20] R. J. Prill, D. Marbach, J. Saez-Rodriguez, P. K. Sorger, L. G. Alexopoulos, X. Xue, N. D. Clarke, G. Altan-Bonnet, and G. Stolovitzky. Towards a rigorous assessment of systems biology models: the DREAM3 challenges. *PLoS ONE*, 5(2):e9202, 2010.

[21] T. Shimamura, S. Imoto, R. Yamaguchi, A. Fujita, M. Nagasaki, and S. Miyano. Recursive regularization for inferring gene networks from time-course gene expression profiles. *BMC Systems Biology*, 3, 2009.

[22] M. Sokolova, N. Japkowicz, and S. Szpakowicz. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Proceedings of the AAAI'06 workshop on Evaluation Methods for Machine Learning*, 2006.

[23] G. D. Tourassi, E. D. Frederick, M. K. Markey, and Jr. C. E. Floyd. Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Medical Physics*, 28(12):2394–2402, 2001.

[24] M. J. van de Vijver, Y. D. He, L. J. van't Veer, H. Dai, A. A. Hart, D. W. Voskuil, G. J. Schreiber, J. L. Peterse, C. Roberts, M. J. Marton, M. Parrish, D. Atsma, A. Witteveen, A. Glas, L. Delahaye, T. van der Velde T, H. Bartelink H, S. Rodenhuis, E. T. Rutgers, S. H. Friend, and R. Bernards. A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medecine*, 347, 2002.

[25] E. P. van Someren, L. F. A. Wessels, E. Backer, and M. J. T. Reinders. Genetic network modeling. *Pharmacogenomics*, 3(4):507–525, 2002.

[26] L. J. van 't Veer, H. Dai, M. J. van de Vijver, Y. D. He, A. A. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsley, R. Bernards, and S. H. Friend. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 406, 2002.

[27] T. Van den Bulcke, K. Van Leemput, B. Naudts, P. van Remortel, H. Ma, A. Verschoren, B. De Moor, and K. Marchal. Syntren: a generator of synthetic gene expression data for design and analysis of structure learning algorithms. *BMC Bioinformatics*, 7(1):43, 2006.

[28] Y. Yang and G. I. Webb. Discretization for naive-bayes learning: managing discretization bias and variance. Technical Report 2003/131 School of Computer Science and Software Engineering, Monash University, 2003.

[29] N. V. Zhukov and S. A. Tjulandin. Targeted therapy in the treatment of solid tumors: Practice contradicts theory. *Biochemistry Moscow*, 2008.