BMC Genomics

RESEARCH

Comparing the reconstruction of regulatory pathways with distinct Bayesian networks inference methods

Adriano V Werhli

Abstract

Background: Inference of biological networks has become an important tool in Systems Biology. Nowadays it is becoming clearer that the complexity of organisms is more related with the organization of its components in networks rather than with the individual behaviour of the components. Among various approaches for inferring networks, Bayesian Networks are very attractive due to their probabilistic nature and flexibility to incorporate interventions and extra sources of information. Recently various attempts to infer networks with different Bayesian Networks approaches were pursued. The specific interest in this paper is to compare the performance of three different inference approaches: Bayesian Networks without any modification; Bayesian Networks modified to take into account specific interventions produced during data collection; and a probabilistic hierarchical model that allows the inclusion of extra knowledge in the inference of Bayesian Networks. The inference is performed in three different types of data: (i) synthetic data obtained from a Gaussian distribution, (ii) synthetic data simulated with Netbuilder and (iii) Real data obtained in flow cytometry experiments.

Results: Bayesian Networks with interventions and Bayesian Networks with inclusion of extra knowledge outperform simple Bayesian Networks in all data sets when considering the reconstruction accuracy and taking the edge directions into account. In the Real data the increase in accuracy is also observed when not taking the edge directions into account.

Conclusions: Although it comes with a small extra computational cost the use of more refined Bayesian network models is justified. Both the inclusion of extra knowledge and the use of interventions have outperformed the simple Bayesian network model in simulated and Real data sets. Also, if the source of extra knowledge used in the inference is not reliable the inferred network is not deteriorated. If the extra knowledge has a good agreement with the data there is no significant difference in using the Bayesian networks with interventions or Bayesian networks with the extra knowledge.

Background

The rapid increase in the availability and diversity of molecular biology data has enabled many discoveries and advances in different fields related with systems biology. Many of these studies were based in a single biological entity or the union of several such entities. Nowadays the research community is realizing that the complexity of an organism is related with the network of single entities rather than with the individual biological entity. It is now clearer that the joint acting of several components through a network of interactions plays a pivotal role in determining the development and sustainability of an organism. Therefore, the study of biological networks is highly relevant. The problem is that these intricate biological networks are mainly unknown. Since we have at our disposal

Correspondence: adrianowerhli@furg.br
Centro de Ciências Computacionais - C3, Universidade Federal do Rio Grande - FURG, RS, Brazil

many different types of measurements taken from the components of these networks one interesting approach would be to try to reconstruct such networks.

In the last few years, several methods for the reconstruction of regulatory networks and biochemical pathways from data have been proposed. These methods were reviewed for example in [1,2].

Differential Equations are the most refined mathematical method to describe biophysical processes. They can describe, for example, the intra-cellular processes of transcription factor binding, diffusion, and RNA degradation; see, for instance, [3]. Such detailed descriptions of the dynamics are essential to an accurate understanding of regulatory networks but they require substantial prior knowledge about the system under investigation. For instance it is necessary to specify how the entities of the system relate with each other and all the parameters of the biochemical reactions. Although differential equations are the most accurate way of representing regulatory networks their use is limited by the necessity of substantial prior knowledge about the system they are representing. At the other extreme is the coarse grain approach of clustering which has been widely applied to the analysis of microarray gene expression data [2,4]. Clustering methods have very low computational costs to extract qualitative information about co-expression, but they are not powerful enough to provide the inference of the detailed structure of the underlying biochemical signalling pathways.

A promising compromise between these two extremes are Machine Learning methods that allow interactions between the nodes in the network to be represented in an abstract way - without the level of detail of the underlying pathways described by Differential Equation models - and to infer these interactions from data in a systems context, that is, distinguishing direct interactions from indirect interactions that are mediated by other nodes in the domain.

A non exhaustive list of methods used to infer the structure of networks from data includes: a system of Coupled Differential Equations [3], Graphical Gaussian Models [5,6], Relevance Networks [7] and Bayesian Networks [8-10]. See [11] for a comparison of some of the methods aforementioned. The main focus in this study is the systematic comparison amongst different BNs approaches. In all the approaches investigated in this paper a score-based inference scheme is followed. In this scheme a score is assigned to a particular model (network structure) given some observed data. The approaches investigated are: (i) a simple BN, (ii) a BN which benefits from interventions that are made in the system of interest during the measurement of the data and (iii) a probabilistic model which enables the use of extra knowledge in the inference. Hereafter we name BNs with interventions as BN-I and BNs with extra knowledge as BN-E.

Results

Evaluation criteria

Not all of the edge directions in a Bayesian network can always be inferred. This is due to the existence of equivalence classes of networks [12] which may lead to partially directed graphs. In view of the presence of directed and partially directed graphs, we apply two different criteria to assess the performance of the methods. In one of the approaches the information about the edge directions is completely discarded. Whenever two nodes are connected by a directed edge this edge is replaced by an undirected one. This approach is called the undirected graph evaluation (UGE). The other approach considers a predicted undirected edge as the superposition of two directed edges, pointing in opposite directions. This approach is called the directed graph evaluation (DGE). The result of the MCMC simulation is a sample of network structures which leads to a matrix of marginal posterior probabilities associated with the edges in a network. This defines a ranking of the edges. This ranking defines a receiver operator characteristics (ROC) curve, where the relative number of true-positive (TP) edges is plotted against the relative number of false-positive (FP) edges. Ideally we would compare the whole ROC curves but this is impractical. Therefore, we use the area under the ROC curve (AUC). The AUC summarizes the results for all the thresholds. A perfect predictor would produce an AUC value of 1.00. Conversely, a random predictor would produce an AUC value around 0.50. In general, bigger area values represent better predictors.

Inference results

MCMC simulations are performed for all the approaches and data sets twice in order to check convergence. The convergence is verified by plotting the posterior probabilities of the edges from two different simulations initializations and checking if the results are similar. Note that this is a necessary but not a sufficient condition for convergence. All the MCMC simulations are executed with 5×10^5 steps from which the first half were discarded as burn-in.

The extra knowledge used in conjunction with the real data in the BN-E approach was obtained from Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways database [13-15] as described in [16]. For the synthetic data sets, obtained from a Gaussian distribution and from Netbuilder, we used two distinct sources of extra-knowledge, one completely correct ($\mathcal{B}_{100}$), and one half correct i.e. half of the entries are correct and the other half is wrong, ($\mathcal{B}_{50}$).

In Figure 1 we present a typical result of the hyperparameters for the two different sources of extra-knowledge used with synthetic data, $\mathcal{B}_{50}$ and $\mathcal{B}_{100}$, and for Real data, KEGG. The results presented are the
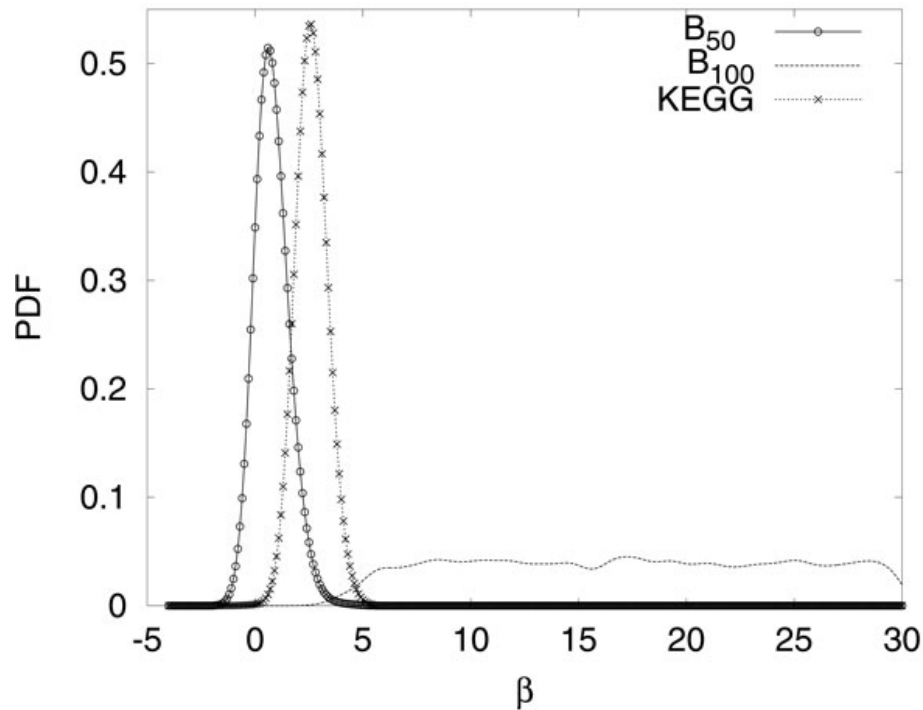
Figure 1 Posterior distribution of hyper-parameter. In the figure $\mathcal{B}_{50}$ is the solid line with circles, $\mathcal{B}_{100}$ is the dashed line and KEGG is the dotted line with crosses. This figure summarizes the typical hyper-parameter found by the BN-E for both sources of extra knowledge available, $\mathcal{B}_{50}$ and $\mathcal{B}_{100}$, when applied to the Gaussian and Netbuilder data and for the extra knowledge from KEGG applied to the Real data. The posterior distribution of the sampled hyper-parameters is estimated with a kernel estimator.

posterior distribution of the hyper-parameters obtained with a kernel estimator. In other words, the hyper-parameter is sampled with an MCMC scheme and in Figure 1 it is shown the distribution of these sampled hyper-parameters. The values of the hyper-parameter are an indicator of how much the extra knowledge agrees with the available data. The completely correct extra-knowledge, $\mathcal{B}_{100}$, is fully integrated in the inference as is indicated by the wide range of large values sampled to the hyper-parameter. Conversely, when considering the half correct extra-knowledge, $\mathcal{B}_{50}$, the low values of the sampled hyper-parameter indicates that the extra-knowledge available is not completely integrated in the inference. For the Real data we can see that the extra knowledge from KEGG is not in complete agreement as is $\mathcal{B}_{100}$ but it presents a slightly higher agreement in terms of the hyper-parameter than the one presented when considering $\mathcal{B}_{50}$.
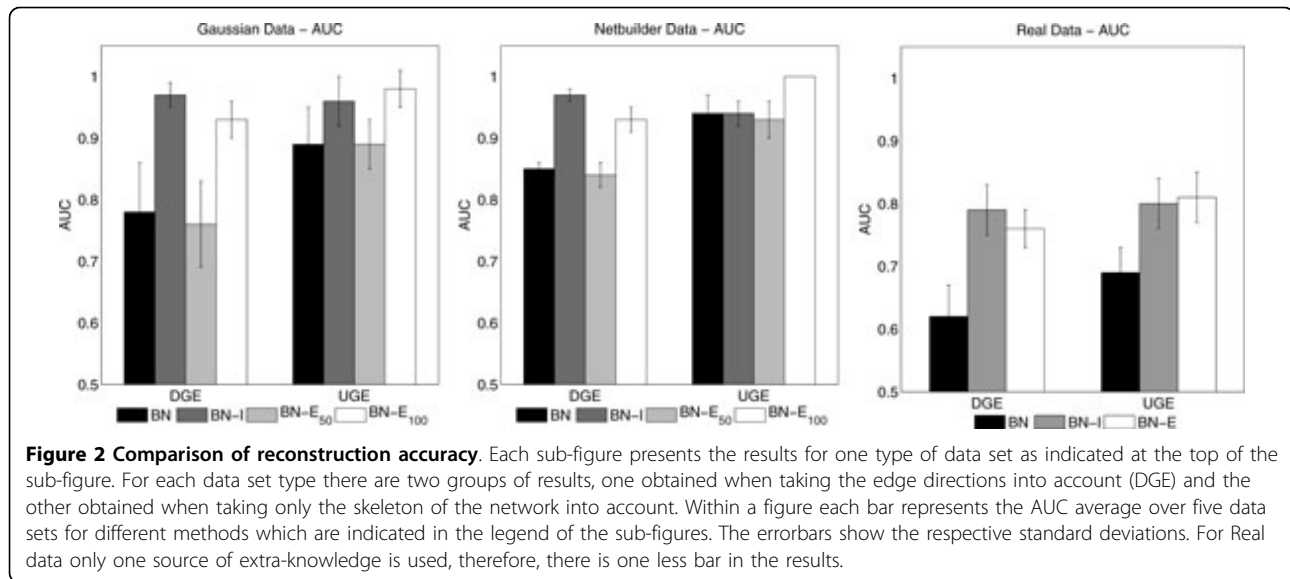
In Figure 2 it is presented a comparison of reconstruction accuracy. Each of the three sub-figure presents the results for one type of data set, Gaussian, Netbuilder and Real data. For each data set type there are two groups of results, one obtained when taking the edge directions into account (DGE) and the other obtained when taking only the skeleton of the network into account. Each bar represents the AUC value, averaged over five data sets, for the different methods as indicated in sub-figures legend. The errorbar shows the respective standard deviations. For Real data only one source of extra-knowledge is used, therefore, there is one less bar in the results.

Discussion

Figure 1 provides an indication about how the method BN-E can benefit from different types of extra knowledge. We can observe that when the method is associated with the extra knowledge, $\mathcal{B}_{100}$, the extra knowledge is effectively used as is indicated by the wide distribution of the sampled hyper-parameters. Conversely, when the method is associated with extra knowledge that is not in perfect agreement with the data, as is the case in $\mathcal{B}_{50}$ and KEGG, it moderately uses the extra knowledge as is evidenced by the distribution of their sampled hyper-parameters close to zero.

One interesting aspect observed is the behaviour of the hyper-parameter of the BN-E approach and the reconstruction accuracy obtained with this method. As we can see from Figure 1 the sampled hyper-parameter for the extra knowledge $\mathcal{B}_{50}$ is much smaller than the one sampled for $\mathcal{B}_{100}$. It is important to notice that this

Figure 2 Comparison of reconstruction accuracy. Each sub-figure presents the results for one type of data set as indicated at the top of the sub-figure. For each data set type there are two groups of results, one obtained when taking the edge directions into account (DGE) and the other obtained when taking only the skeleton of the network into account. Within a figure each bar represents the AUC average over five data sets for different methods which are indicated in the legend of the sub-figures. The errorbars show the respective standard deviations. For Real data only one source of extra-knowledge is used, therefore, there is one less bar in the results.

is the expected behaviour of the hyper-parameter. If we observe Figures 2 left and center we can see that the accuracy of the recovered structure is much higher for $\mathcal{B}_{100}$ than for $\mathcal{B}_{50}$ and that this difference is more pronounced for the DGE evaluation criterion. Also, note that the AUC values obtained for $\mathcal{B}_{50}$ are in fact very similar to the ones obtained by the BN approach. This behaviour suggests that in order to the BN-E approach to take advantage from the extra knowledge available it has to have a very good agreement with the data. Moreover, if the data and the extra knowledge are not in agreement i.e. the extra knowledge is not beneficial to the inference, the BN-E approach does not suffer any impairment.

Observing the results for synthetic data in Figures 2 left and central we see that the methods BN-I and BN-E$_{100}$ clearly outperform the others specially when considering the edge directions (DGE criterion). This suggests that the increase in the accuracy of the recovered networks is related with the edge directions i.e. these methods provide a way to the break up the symmetries which imply in the equivalence classes. It is also possible to note that the addition of the half correct extra knowledge, $\mathcal{B}_{50}$, did not improve the results obtained with the BN approach. It is important to emphasize that the extra knowledge $\mathcal{B}_{50}$ has half of its entries correct and the other half incorrect. As there is only one hyper-parameter associated with the extra knowledge if this hyper-parameter increases the inference method would use wrong information and certainly the result would be worse than without any extra knowledge. Therefore, this explains why the BN-E$_{50}$ approach had the hyper-parameter sampled at very small values and, hence, has not improved the accuracy of the reconstructed network.

In Figure 2 right we see that both BN-I and BN-E outperformed BN increasing substantially the accuracy of the recovered networks for the Real data. Interestingly though, in this case both the DGE and UGE criteria have benefited of the extra information introduced by both approaches. We see that the hyper-parameter associated with the KEGG prior is slightly higher than the hyper-parameter associated with the $\mathcal{B}_{50}$. This is explained by the construction of the KEGG prior where unknown entries were regarded as unknown ($\mathcal{B}_{i,j} = 0.5$) as opposite to the extra knowledge $\mathcal{B}_{50}$ where half of the entries are wrong.

Conclusion

BNs are very attractive to the inference of the structure of networks by various reasons. One of the main advantages of BNs is its flexibility. In this paper we compared different BNs approaches where two of them are extensions of the classical BNs framework. The essence of both of these extensions of BNs is the inclusion of knowledge other than the data in the inference. If the BN-I interventions are taken into consideration and in the BN-E extra knowledge is added to the learning scheme.

Observing the results in Figure 2 we can conclude that both BN-I and BN-E$_{100}$ perform better than the simple BN. This performance is significantly better when the comparison takes into account the edge directions (DGE score). This leads to the conclusion that both methods in fact perform better because they are able to destroy the equivalence classes symmetries. Another interesting conclusion is obtained when we observe Figure 1 in conjunction with Figure 2. It is clear that in order to the BN-E to benefit from the available extra-knowledge this has to

have a very good agreement with the data as is the case with the $\mathcal{B}_{100}$ extra knowledge. If this is not the case, the behaviour is similar to the one presented by the extra knowledge $\mathcal{B}_{50}$ where the hyper-parameter is sampled at values close to zero effectively switching off the use of the extra knowledge, hence not improving the AUC scores.

Interestingly there are no significant differences when comparing the two best methods, BN-I and BN-E$_{100}$, as can be observed in Figure 2. Both performed equally well in all the data sets used in this work. Therefore, there is no clear indication about which of these methods should be used. Nonetheless, it is clear that using any of them is a great advantage over the simple BN. The decision to chose among BN-E and BN-I will have to be made according to the system in study and the data available. For systems in which there are plenty of available extra knowledge this study suggests that it is not necessary to perform such expensive experiments with interventions. Conversely, if the system under scrutiny does not have extra knowledge available, it may be advisable to perform experiments with interventions. It is worth to note that even if there is plenty of extra knowledge available it is impossible to know beforehand if this will be in good agreement with the data. One indication about the quality of the extra knowledge can be obtained by the observation of the distribution of the sampled hyper-parameters as it is presented in Figure 1.

The main conclusion is that the use of more refined Bayesian network models significantly improves the results. Both more refined methods, BN-E and BN-I, performed equally well and, hence, their choice should be made according to the quality and availability of the data obtained from the system under investigation.

Methods
Bayesian Networks - BNs
Bayesian Networks (BNs) are a combination of probability theory and graph theory. A graphical structure $\mathcal{M}$, a family of conditional probability distributions $\mathcal{F}$ and their parameters $\mathbf{q}$, fully specify a BN. The graphical structure $\mathcal{M}$ of a BN consists of a set of nodes and a set of directed edges. The nodes represent random variables, while the edges indicate conditional dependence relations. The family of conditional probability distributions $\mathcal{F}$ and their parameters $\mathbf{q}$ specify the functional form of the conditional probabilities associated with the edges, that is, they indicate the nature of the interactions between nodes and the intensity of these interactions. A BN is characterized by a simple and unique rule for expanding the joint probability in terms of simpler conditional probabilities. This follows the local Markov property: *A node is conditionally independent of its non descendants given its parents.* Due to this property, the structure $\mathcal{M}$ of a BN has necessarily to be a directed acyclic graph (DAG), that is, a network without any directed cycles. Let $X_1, X_2, ..., X_N$ be a set of random

variables represented by the nodes $i \in \{1, ..., N\}$ in the graph, define $\pi_i[G]$ to be the parents of node X_i in graph G, and let $X_{\pi_i}[G]$ represent the set of random variables associated with $\pi_i[G]$. Then we can write the expansion for the joint probability as $P(X_1, ..., X_N) = \prod_{i=1}^{N} P(X_i|X_{\pi_i}[G])$.

The task of learning a BN structure in a score-based approach consists in devising a BN structure from a given set of training data $\mathcal{D}$. The main aim is to find a DAG structure that better explains the data available for learning. If we define that $\mathbb{M}$ is the space of all models, the first goal is to find a model $\mathcal{M}^* \in \mathbb{M}$ that is most supported by the data $\mathcal{D}$, $\mathcal{M}* = \arg \max_{\mathcal{M}} \{P(\mathcal{M}|\mathcal{D})\}$. Having the best structure $\mathcal{M}^*$ and the data $\mathcal{D}$, we can now find the best parameters, $\mathbf{q} = \arg \max_{\mathbf{q}} \{P(\mathbf{q}|\mathcal{M}^*, \mathcal{D}|\}$ If we apply Bayes' rule we get $P(\mathcal{M}|\mathcal{D}) \propto P(\mathcal{D}|\mathcal{M})P(\mathcal{M})$ where the marginal likelihood implies an integration over the whole parameter space:

$$P(\mathcal{D}|\mathcal{M}) = \int P(\mathcal{D}|\mathbf{q}, \mathcal{M})P(\mathbf{q}|\mathcal{M})d\mathbf{q} \qquad (1)$$

The integral in Equation 1, our score, is analytically tractable when the data is complete and the prior $P(\mathbf{q}|\mathcal{M})$ and the likelihood $P(\mathcal{D}|\mathbf{q}, \mathcal{M})$ satisfies certain regularity conditions [17,18].

According to Equation 1 we have a way to assign a score to a graphical structure given a data set. However, the search for high scoring structures is not trivial. It is impossible to list the whole set of structures because its number increases super-exponentially with the number of nodes. Also when considering an sparse data set $P(\mathcal{M}|\mathcal{D})$ is diffuse, meaning that $P(\mathcal{M}|\mathcal{D})$ will not be properly represented by a single structure $\mathcal{M}^*$. Hence, a Markov chain Monte Carlo (MCMC) scheme is adopted [19], which under fairly general regularity conditions is theoretically guaranteed to converge to the posterior distribution [20]. Given a network structure $\mathcal{M}_{\text{old}}$, a network structure $\mathcal{M}_{\text{new}}$ is proposed from the proposal distribution $\mathcal{Q}(\mathcal{M}_{\text{new}}|\mathcal{M}_{\text{old}})$, which is then accepted according to the standard Metropolis-Hastings [20] scheme with the following acceptance probability:

$$A = \min\left\{\frac{P(\mathcal{D}|\mathcal{M}_{\text{new}})P(\mathcal{M}_{\text{new}})Q(\mathcal{M}_{\text{old}}|\mathcal{M}_{\text{new}})}{P(\mathcal{D}|\mathcal{M}_{\text{old}})P(\mathcal{M}_{\text{old}})Q(\mathcal{M}_{\text{new}})(\mathcal{M}_{\text{old}})}, 1\right\} \quad (2)$$

In this paper we use the standard MCMC proposal which consists in to propose, at each interaction, one of the basic operations of adding, removing or reversing an edge. For more details about this scheme see [10].

Bayesian Networks with Interventions - BN-I
Nowadays molecular biology has different techniques for producing interventions in biological systems, for instance,

knocking genes down with RNA interference or transposon mutagenesis. The consequence is that the components of the system which are targeted by the interventions are no longer subject to the internal dynamics of the system under investigation. The components of the biological system can be either activated (up-regulated) or inhibited (down-regulated) and under this external intervention their values are no longer stochastic. The intervened components are not subject to the internal dynamics of the system, hence their values are deterministic. However, the other components which are not intervened are influenced by these deterministic values. Therefore, interventions are very useful to break the symmetries within the equivalence classes of BNs and consequently to the discovery of putative causal relationships. For a discussion about equivalence classes see [21] and for a discussion about putative causal relationships see [12,22].

In order to incorporate the interventions under the BN framework two small modifications are necessary. The calculation of the score for observational data $P(\mathcal{D}|\mathcal{M})$ as defined in Equation (1) is modified. Effectively the measurements of a node X_i under intervention are removed from the computation of the score.

The second necessary modification is related to the definition of equivalence classes. In [23] it is defined the Transition Sequence equivalent networks (TS-equivalent). Two networks $\mathcal{M}_1$ and $\mathcal{M}_2$ are TS-equivalent if and only if they have the same skeleton, the same set of v-structures and the same set of parents for all manipulated variables. All edges connected with an intervened node become directed when the concept of TS-equivalence is applied. Therefore, new v-structures are formed and further edges become directed. In order to obtain the TS-equivalent DAG the procedure presented by [24] is applied. For each intervened node in the network two dummy nodes are added each with one directed edge pointing from the dummy node to the intervened node. The new DAG now with the dummy nodes added is converted to a CPDAG (for a discussion about CPDAGs see [25]). Finally the dummy nodes are removed and we have the DAG TS-equivalent graph.

The sampling scheme of the BNs-I is the same of the BNs and is given by Equation 2.

Bayesian Networks with addition of Extra knowledge - BN-E

In order to be able to incorporate extra knowledge in the inference of networks it is necessary to define a function that measures the agreement between a given network structure $\mathcal{M}$ and the extra knowledge that we have at our disposal, $\mathcal{B}$. We call this agreement measure 'energy' following the approach proposed in [26].

A network structure $\mathcal{M}$ is represented by an adjacency matrix where each entry $\mathcal{M}_{ij}$ can be either 0 or 1 representing respectively the absence and the presence of an edge between node-i and node-j. The prior knowledge matrix, or belief matrix, $\mathcal{B}$, is defined to have entries $\mathcal{B}_{ij} \in [0, 1]$ representing our knowledge about the node interactions. An entry $\mathcal{B}_{ij} = 0.5$ denotes that we do not have any information about the presence or absence of an edge between node-i and node-j. If $0 \leq \mathcal{B}_{ij} < 0.5$ we have prior evidence that the edge between node-i and node-j is absent and the evidence is stronger as $\mathcal{B}_{ij}$ is closer to 0. At last, if $0.5 < \mathcal{B}_{ij} \leq 1$ we have prior evidence that there is a directed edge pointing from node-i to node-j. The evidence is stronger as $\mathcal{B}_{ij}$ is closer to 1. It is important to note that the entries in our belief matrix are not proper probabilities and they only express our belief, or knowledge obtained from other sources, about the relationships among nodes.

Having defined how to represent a BN structure, $\mathcal{M}$, and the extra belief, $\mathcal{B}$, the energy of a network is defined as:

$$E(\mathcal{M}) = \sum_{i,j=1}^{N} |\mathcal{B}_{i,j} - \mathcal{M}_{i,j}| \tag{3}$$

where N is the total number of nodes in the studied domain. From Equation 3 it is clear that the energy E is zero for a perfect match between the prior knowledge $\mathcal{B}$ and the actual network structure $\mathcal{M}$, while increasing values of E indicate an increasing mismatch between $\mathcal{B}$ and $\mathcal{M}$.

Following the work of [26] we integrate the prior knowledge expressed by Equation 3 into the inference procedure, and define the prior distribution over network structures, $\mathcal{M}$, to take the form of a Gibbs distribution:

$$P(\mathcal{M}|\beta) = \frac{e^{-\beta E(\mathcal{M})}}{\sum_{\mathcal{M} \in \mathbb{M}} e^{-\beta E(\mathcal{M})}} \tag{4}$$

where the energy E($\mathcal{M}$) was defined in Equation 3, β is a hyper-parameter that corresponds to an inverse temperature in statistical physics, and the denominator is a normalizing constant that is usually referred to as the partition function. Note that the summation in the denominator extends over the set of all possible network structures. The hyper-parameter β can be interpreted as a factor that indicates the strength of the influence of the prior knowledge relative to the data. For $\beta \rightarrow 0$, the prior distribution defined in Equation 4 becomes flat and uninformative about the network structure. Conversely, for $\beta \rightarrow \infty$, the prior distribution becomes sharply peaked at the network structure with the lowest energy.

For Dynamic Bayesian Networks the summation in the denominator of Equation 4 can be computed exactly and efficiently as discussed in [16] with

$$Z(\beta) = \prod_n \sum_{\pi_{\mathcal{M}}(n)} e^{-\beta\varepsilon(n,\pi_{\mathcal{M}}(n))}. \tag{5}$$

In this paper we apply the method only to static BNs and thus the summation in the denominator of Equation 4 is in fact an upper bound to the true value. This happens because this summation includes all possible structures and we are only interested in the DAG structures. Furthermore, throughout this paper we use a fan-in restriction of three as has been proposed in several other applications, for instance see [27-29]. This fan-in restriction makes the summation over all structures closer to the summation of only the DAGs as it reduces the number of densely connected networks. The partition function approximation has been investigated elsewhere [16,30] and was not found to pose a problem to the proposed method.

BN-E MCMC sampling scheme

At this point, having the prior probability distribution over network structures defined, an MCMC scheme to sample both the hyper-parameters and the network structures from the posterior distribution $P(\mathcal{M}, \beta | \mathcal{D})$ is proposed.

A new network structure $\mathcal{M}_{\text{new}}$ and a new hyper-parameter β_{new} are proposed respectively from the proposal distributions $\mathcal{Q}(\mathcal{M}_{\text{new}} | \mathcal{M}_{\text{old}})$ and $\mathcal{R}(\beta_{\text{new}} | \beta_{\text{old}})$. This proposed move is then accepted according to the Metropolis-Hastings update rule [20] with the following acceptance probability:

$$A = \min\left\{ \frac{P(\mathcal{D}|\mathcal{M}_{\text{new}})}{P(\mathcal{D}|\mathcal{M}_{\text{old}})} \times \frac{P(\mathcal{M}_{\text{new}}|\beta_{\text{new}})}{P(\mathcal{M}_{\text{old}}|\beta_{\text{old}})} \times \frac{P(\beta_{\text{new}})}{P(\beta_{\text{old}})} \times \frac{Q(\mathcal{M}_{\text{old}}|\mathcal{M}_{\text{new}})}{Q(\mathcal{M}_{\text{new}}|\mathcal{M}_{\text{old}})} \times \frac{R(\beta_{\text{old}}|\beta_{\text{new}})}{R(\beta_{\text{new}}|\beta_{\text{old}})}, 1 \right\} \tag{6}$$

which was expanded following the conditional independence relationships depicted in Figure 3.

In order to increase the acceptance probability which in turn can augment the mixing and convergence of the Markov chain the move is separeted into two submoves. In the first move a new network structure $\mathcal{M}_{\text{new}}$ is sampled from the proposal distribution $Q(\mathcal{M}_{\text{new}}|\mathcal{M}_{\text{old}})$

while keeping the hyper-parameter β fixed. Next, we sample a new hyper-parameter β from the proposal distribution $R(\beta_{\text{new}}|\beta_{\text{old}})$ for a fixed network structure $\mathcal{M}$. The two sub-moves are iterated until a convergence criterion is satisfied.

Data sets

One very interesting aspect when comparing different methods applied to the inference of the structure of networks is the ability to compare how they perform when faced with real data sets. In our case a real data set means data obtained with real experiments from a real biological system. Also, the comparison among the methods with real data is only possible if the network which the data was generated from is known. We call this known network the *gold-standard* network. Taking these considerations into account we use data from flow cytometry experiments obtained by [31] where the Raf signalling pathway, see Figure 4, was studied. This particular data set is very interesting as it provides high quality measurements, large amounts of data, intervened data and a *gold-standard* network.

Because the interest is to compare the BNs approaches in the context of inference of networks, where the data available are usually sparse, we down sampled the original data to 100 data points. Furthermore, we average the results over five data sets. The observational data is obtained from the original data where no interventions were realized. The interventional data is sampled from all the interventions realized in the original data and is composed by: 16 data points without intervention; 14 data points for each of the inhibited proteins (*AKT, PKC, PIP2, MEK*) and 14 data points for each of the activated proteins (*PKC, PKA*) proteins, performing a total of 100 data points.

In order to further investigate how the methods compare synthetic data sets were also prepared. These data are obtained from two different sources: a linear Gaussian distribution and a simulation tool named Netbuilder [32,33]. In both cases, the data is obtained from the known structure of Figure 4.

Figure 3 Probabilistic Model. The probabilistic graphical model represents conditional independence relationships between the data $\mathcal{D}$, the network structure $\mathcal{M}$, and the hyper-parameter of the prior on $\mathcal{M}$. The conditional independence relationships can be obtained from the graph according to the standard rules of factorization in Bayesian networks, as discussed, e.g., in [34]. This leads to the following expansion: $P(\mathcal{D}, \mathcal{M}, \beta) = P(\mathcal{D}|\mathcal{M})P(\mathcal{M}|\beta)P(\beta)$.
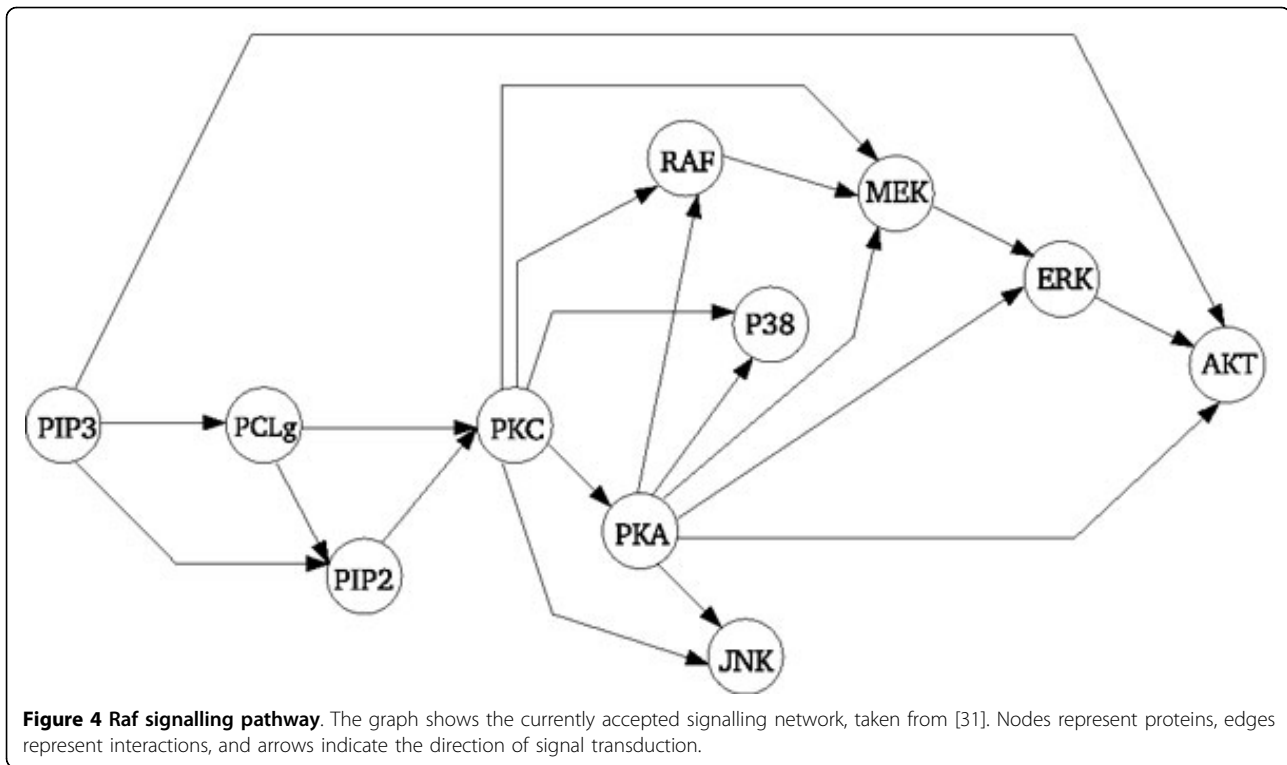
Figure 4 Raf signalling pathway. The graph shows the currently accepted signalling network, taken from [31]. Nodes represent proteins, edges represent interactions, and arrows indicate the direction of signal transduction.

Considering $\mathcal{N}(.)$ to denote the Normal distribution, a random variable x_i is sampled from a linear Gaussian distribution with value distributed according to $x_i \sim \mathcal{N}(\sum_k w_{ik} x_k, \sigma^2)$ where the sum extends over all parents of node i and xk represents the value of node k. The standard deviation is set to $\sigma = 0.1$ and the interaction strengths $|w_{ik}|$ are sampled from the uniform distribution over the interval [0.5, 2], where the sign of w_{ik} is randomly varied.

In Netbuilder a sigma-pi formalism is implemented as an approximation to the solution of a set of Ordinary Differential Equations that models enzyme-substrate reactions, allowing the acquisition of data typical of signals measured in molecular biology. The data sets simulated with Netbuilder are closely related to real data sets when compared with the Gaussian data. For more details about the data generation see [11].

Competing interests
The authors declare that they have no competing interests.

References
1. De Jong H: **Modeling and Simulation of Genetic Regulatory Systems: A Literature Review.** *Journal of Computational Biology* 2002, **9**:67-103.
2. D'haeseleer P, Liang S, Somogyi R: **Genetic network inference: from co-expression clustering to reverse engineering.** *Bioinformatics* 2000, **16(8)**:707-726.
3. Chen T, He HL, Church GM: **Modeling gene expression with differential equations.** *Pacific Symposium on Biocomputing* 1999, **4**:29-40.
4. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster Analysis and Display of Genome-Wide Expression Patterns.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**:14863-14868.
5. Schäfer J, Strimmer K: **An Empirical Bayes Approach to Inferring Large Graphical Gaussian Models from Microarray Data.** *Bioinformatics* 2004, **1**:1-13.
6. Opgen-Rhein R, Strimmer K: **Inferring gene dependency networks from genomic longitudinal data: a functional data approach.** *REVSTAT* 2006, **4**:53-65.
7. Butte AJ, Kohane IS: **Mutual information relevance networks: functional genomic clustering using-pairwise entropy measurements.** *Pac Symp Biocomput* 2000, 418-429.
8. Friedman N, Nachman I, Pe'er D: **Learning Bayesian Network Structure from Massive Datasets: The "Sparse Candidate" Algorithm.** *Proceedings of the Fifteenth Annual Conference on Uncertainty in Artificial Intelligence* San Francisco, CA: Morgan Kaufmann Publishers; 1999, 196-205.
9. Friedman N, Linial M, Nachman I, Pe'er D: **Using Bayesian Networks to Analyze Expression Data.** *Journal of Computational Biology* 2000, **7**:601-620.
10. Husmeier D, Dybowski R, Roberts S: *Probabilistic Modeling in Bioinformatics and Medical Informatics* Advanced Information and Knowledge Processing, New York: Springer; 2005.
11. Werhli AV, Grzegorczyk M, Husmeier D: **Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical Gaussian models and Bayesian networks.** *Bioinformatics* 2006, **22(20)**:2523-2531.
12. Pearl J: *Causality: Models, Reasoning and Intelligent Systems* London, UK: Cambridge University Press; 2000.

13. Kanehisa M: **A database for post-genome analysis.** *Trends Genet* 1997, **13**:375-376.

14. Kanehisa M, Goto S: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Research* 2000, **28**:27-30.

15. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita K, Kawashima MIS, Katayama T, Araki M, Hirakawa M: **From genomics to chemical genomics: new developments in KEGG.** *Nucleic Acids Research* 2006, **34**:D354-357.

16. Werhli AV, Husmeier D: **Reconstructing Gene Regulatory Networks with Bayesian Networks by Combining Expression Data with Multiple Sources of Prior Knowledge.** *Statistical Applications in Genetics and Molecular Biology* 2007, **6**:Article 15.

17. Heckerman D: **Learning Gaussian Networks.** Tech Rep MSR-TR-94-10, Microsoft Research, Redmond, Washington; 1994.

18. Heckerman D: *A tutorial on learning with Bayesian networks* Tech Rep MSR-TR-95-06, Microsoft Research, Redmond, Washington; 1995.

19. Madigan D, York J: **Bayesian graphical models for discrete data.** *International Statistical Review* 1995, **63**:215-232.

20. Hastings WK: **Monte Carlo sampling methods using Markov chains and their applications.** *Biometrika* 1970, **57**:97-109.

21. Verma T, Pearl J: **Equivalence and Synthesis of CausalModels.** *Proceedings of the 6th Annual Conference on Uncertainty in Artificial Intelligence(UAI-91)* New York, NY: Elsevier Science; 1991, 255-270.

22. Cooper G, Glymour C: *Computation, Causation, and Discovery* Cambridge, MA: MIT Press; 1999.

23. Tian J, Pearl J: *Causal discovery from changes: a Bayesian approach* 2001.

24. Wernisch L, Pournara I: **Reconstruction of gene networks using Bayesian learning and manipulation experiments.** *Bioinformatics* 2004, **20**:2934-2942.

25. Chickering DM: **Learning Equivalence Classes of Bayesian-Network Structures.** *Journal of Machine Learning Research* 2002, **2**:445-498.

26. Imoto S, Higuchi T, Goto T, Tashiro K, Kuhara S, Miyano S: **Combining Microarrays and Biological Knowledge for Estimating Gene Networks via Bayesian Networks.** *Proceedings IEEE Computer Society Bioinformatics Conference* 2003, , CSB'03: 104-113.

27. Friedman N, Linial M, Nachman I, Pe'er D: **Using Bayesian Networks to Analyze Expression Data.** *Journal of Computational Biology* 2000, **7**(3/4):601-620.

28. Friedman N, Koller D: **Being Bayesian about Network Structure.** *Machine Learning* 2003, **50**:95-126.

29. Husmeier D: **Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks.** *Bioinformatics* 2003, **19**:2271-2282.

30. Werhli AV, Husmeier D: **Gene Regulatory Network Reconstruction by Bayesian Integration of Prior Knowledge and/or Different Experimental Conditions.** *Journal of Bioinformatics and Computational Biology* 2008, **6**:543-572.

31. Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP: **Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data.** *Science* 2005, **308**(5721):523-529.

32. Yuh CH, Bolouri H, Davidson EH: **Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene.** *Science* 1998, **279**:1896-1902.

33. Yuh CH, Bolouri H, Davidson EH: **Cis-regulatory logic in the endo16 gene: switching from a specification to a differentiation mode of control.** *Development* 2001, **128**:617-629.

34. Heckerman D: **A Tutorial on Learning with Bayesian Networks.** In *Learning in Graphical Models, Adaptive Computation and Machine Learning.* Cambridge, Massachusetts: MIT Press;Jordan MI 1999:301-354.