USING DIRECTED INFORMATION FOR INFLUENCE DISCOVERY IN INTERCONNECTED DYNAMICAL SYSTEMS

Arvind Rao [a,b,c], Alfred O. Hero [a,b], David J. States [b] and James Douglas Engel [c]

[a]Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA
[b]Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA
[c]Cell and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA

ABSTRACT

Structure discovery in non-linear dynamical systems is an important and challenging problem that arises in various applications such as computational neuroscience, econometrics, and biological network discovery. Each of these systems have multiple interacting variables and the key problem is the inference of the underlying structure of the systems (which variables are connected to which others) based on the output observations (such as multiple time trajectories of the variables).

Since such applications demand the inference of directed relationships among variables in these non-linear systems, current methods that have a linear assumption on structure or yield undirected variable dependencies are insufficient. Hence, in this work, we present a methodology for structure discovery using an information-theoretic metric called directed time information (DTI). Using both synthetic dynamical systems as well as true biological datasets (kidney development and T-cell data), we demonstrate the utility of DTI in such problems.

Keywords: Mutual Information; Directed Information; transcription regulatory network.

1. INTRODUCTION

Estimating the structure of dynamical systems is an interesting and well-studied problem. Ranging from parameter estimation in linear and non-linear dynamical systems to inference of interacting variables over graph-topologies, the various aspects of structure discovery has several applications. Due to recent interest in areas like gene regulatory network inference, there is renewed interest to use principled metrics that can aid structure discovery in these scenarios from multiple realizations of system trajectories.

Consider, for example, a dynamical system with K state variables $\underline{\mathbf{g}} = [g_1, g_2, \ldots, g_K]^T$. $\underline{\mathbf{g}}$ is the state vector and the time evolution of a simple non-linear system can be written as $\underline{\mathbf{g}}_{t+1} = A \times f(\underline{\mathbf{g}}_t) + \epsilon_t$, where A is a transition matrix, and $f(.)$ is some non-linear transformation of the state vector, ϵ_t is the noise term.

As an example, consider the following dynamical system evolution equations:

Further author information: (Send correspondence to Arvind Rao)
Arvind Rao: E-mail: ukarvind@umich.edu

$$
\begin{aligned}
g_{2,t} &= \frac{1}{2} g_{1,t-1} + \frac{1}{3} g_{3,t-2} + g_{7,t-1}; \\
g_{4,t} &= g_{2,t-1}^2 + g_{3,t-1}^{1/2}; \\
g_{5,t} &= g_{2,t-2} + g_{4,t-1}; \\
g_{6,t} &= g_{4,t-1} + g_{2,t-2}^{1/2}; \\
g_{7,t} &= \frac{1}{2} g_{4,t-1}^{1/3}; \\
g_{8,t} &= \frac{1}{2} g_{6,t-1}^{1/3} + \frac{1}{3} g_{7,t-1}^{1/2}; \\
g_{9,t} &= \frac{2}{3} g_{4,t-1}^{2/3} + \frac{1}{4} g_{7,t-2}^{1/2};
\end{aligned}
$$

In these equations, both non-linearities and lagged relationships amongst interacting variables are accounted for. The main question for this work is: given *multiple realizations* of the individual time trajectories of these nine variables, can we infer directed network amongst these variables, where each directed link $g_i \rightarrow g_j$ represents the time-level influence of an effector variable g_i on the variable g_j. This question can be resolved in two directions:

1. Can we come up with a viable influence metric for inference of directed dependence among variables?
2. Can we use this metric for *both* supervised and unsupervised network inference?
    - *Unsupervised Network Inference*: In this part a directed dependency graph is inferred, using no *apriori* information about possible effector variables. Such a procedure explores all $G(G-1)$, interactions to build the directed graph ($G = 9$).
    - *Supervised Network Inference*: In this component, we will find effectors for a variable of interest using only a restricted subset of variables. This answers a question of the type: "Which variables influence the variable g_7"?

The motivation for the set-up above is the resolution of directed graphs in gene regulatory networks. In computational biology, the problem of network inference among genes has received considerable interest (Rangel,[38] Beal[5]). Using time series expression data, available from microarray experiments, biologists are interested to discover gene dependencies and their meaning in the context of biological processes such as transcriptional regulation. Since in these biochemical reactions, the notion of lag and non-linear transcriptional kinetics is closely related to biological processes, our setup in the equations above closely mimics such scenarios.

Other applications where the use of an information flow metric becomes useful is in computational neuroscience (Hartemink[20]) and econometrics (Geweke[17]). In neuroscience it is useful to determine directed dependencies among various brain regions based on time series data from electrodes embedded in these regions. In econometrics, there is an interest to recover dependencies among various time series related to an economic phenomenon (e.g.: relation between economic output (GDP) and inflation over a 10 year time period). We note that in the case of large time series data (wherein the sampling interval is much shorter compared to the length of the observation), several metrics have been proposed in the domain of neuroscience, such as directed transinformation (Williams[42]). These metrics have been extremely useful in such contexts, though an extension to short time series with larger time intervals between sampling instants would be very useful. In this work however, we will focus on applications from computational biology to illustrate our methods for such scenarios.

2. DTI FORMULATION

As alluded to above, there is a need for a viable influence metric that can find relationships between the "effector" variable (g_i) and the target variable (g_j). Several such metrics have been proposed – both generally and in the

context of biological networks, such as correlation, coefficient of determination (CoD), mutual information etc. To alleviate the challenge of detecting non-linear variable interactions, an information theoretic measure like mutual information has been used to infer the conditional dependence among variables by exploring the structure of the joint distribution of the variable expression profiles (Califano *et.al*[30]). However, the absence of a directed dependence metric has hindered the utilization of the full potential of information theory. In this work, we examine the applicability of one such metric – the directed time information criterion (DTI), for the inference of non-linear, directed variable influences.

The DTI, which is a measure of the directed dependence between two N-length random processes $X \equiv X^N$ and $Y \equiv Y^N$, is given by Massey[32] :

$$I(X^N \to Y^N) = \sum_{n=1}^{N} I(X^n; Y_n | Y^{n-1}) \tag{1}$$

Here, Y^n denotes $(Y_1, Y_2, .., Y_n)$, i.e., a segment of the realization of a random process Y and $I(X^N; Y^N)$ is the Shannon mutual information (Cover & Thomas[12]).

An interpretation of the above formulation for DTI is in order. To infer the notion of influence between two time series (mRNA expression data) we find the mutual information between the entire evolution of variable X (up to the current instant n) and the current instant of Y (Y_n), given the evolution of variable Y up to the previous instant $n-1$ (i.e. Y^{n-1}). This is done for every instant, $n \in (1, 2, \ldots, N)$, in the N - length expression time series.

As already known, $I(X^N; Y^N) = H(X^N) - H(X^N|Y^N)$, with $H(X^N)$ and $H(X^N|Y^N)$ being the entropy of X^N and the conditional entropy of X^N given Y^N, respectively. Using this definition of mutual information, the DTI can be expressed in terms of individual and joint entropies of X^N and Y^N. The task of N-dimensional entropy estimation is an important one and due to computational complexity and moderate sample size, histogram estimation of multivariate density is unviable. However, several methods exist for consistent entropy estimation of multivariate small sample data (LearnedMiller[25] , Nemenman[33] , Paninski[36] , Wilett[50]). In the context of microarray expression data, wherein probe-level and technical/biological replicates are available, we use the method of Learned-Miller[25] for entropy estimation.

From (1), we have,

$$I(X^N \to Y^N) = \sum_{n=1}^{N} [H(X^n|Y^{n-1}) - H(X^n|Y^n)] = \sum_{n=1}^{N} \{[H(X^n, Y^{n-1}) - H(Y^{n-1})] - [H(X^n, Y^n) - H(Y^n)]\} \tag{2}$$

- To evaluate the DTI expression in *Eqn*.2, we need to estimate the entropy terms $H(X^n, Y^{n-1})$, $H(Y^{n-1})$, $H(X^n, Y^n)$ and $H(Y^n)$. This involves the estimation of marginal and joint entropies of n random variables, each of which are R dimensional, R being the number of replicate realizations.

- Though some approaches need the estimation of probability density of the R-dimensional multivariate data (X^n) prior to entropy estimation, one way to circumvent this is to the use the method proposed in Learned-Miller[25] . This approach uses a Voronoi tessellation of the R-dimensional space to build nearly uniform partitions (of equal mass) of the density. The set of Voronoi regions $(V^1, V^2, \ldots, V^n)$ for each of the n points in R-dimensional space is formed by associating with each point X_k, a set of points V^k that are closer to X_k than any other point X_l, where the subscripts k and l pertain to the k^{th} and l^{th} time instants of variable expression.

- Thus, the entropy estimator is expressed as : $\hat{H}(X^n) = \frac{1}{n}\sum_{i=1}^{n} \log(nA(V^i))$, where $A(V^i)$ is the R-dimensional volume of Voronoi region V^i. $A(V^i)$ is computed as the area of the polygon formed by the vertices of the convex hull of the Voronoi region V^i. This estimate has low variance and is asymptotically efficient[26] .

To obtain the DTI between any two variables of interest (X and Y) with N-length expression profiles X^N and Y^N respectively, we plug in the entropy estimates computed above into the above expression (2).

From the definition of DTI, we know that $0 \leq I(X_i^N \rightarrow Y^N) \leq I(X_i^N; Y^N) < \infty$. For easy comparison with other metrics, we use a normalized DTI metric (see Appendix) given by $\rho_{DTI} = \sqrt{1 - e^{-2I(X^N \rightarrow Y^N)}} = \sqrt{1 - e^{-2\sum_{i=1}^{N} I(X^i; Y_i | Y^{i-1})}}$. This maps the large range of DTI, ($[0, \infty]$) to lie in $[0, 1]$. Another point of consideration is to estimate the significance of the 'true' DTI value compared to a null distribution on the DTI value (i.e. what is the chance of finding the DTI value by chance from the series X and Y). This is done using empirical p-value estimation after bootstrap resampling (Sec: 3). A threshold p-value of 0.05 is used to estimate the significance of the true DTI value in conjunction with the the density of a random data permutation, as outlined below.

3. SIGNIFICANCE ESTIMATION OF DTI

We now outline a procedure to estimate the empirical p-value to ascertain the significance of the normalized directed information $\hat{I}(X^N \rightarrow Y^N)$ between any two N-length time series $X \equiv X^N = (X_1, X_2, \ldots, X_N)$, and $Y \equiv Y^N = (Y_1, Y_2, \ldots, Y_N)$. In our case, the detection statistic is $\Theta = \hat{I}(X^N \rightarrow Y^N)$ and the chosen acceptable p-value is α.

The overall bootstrap based test procedure is (Tibshirani *et.al*[15] , Silverman[37] , Polland[2]):

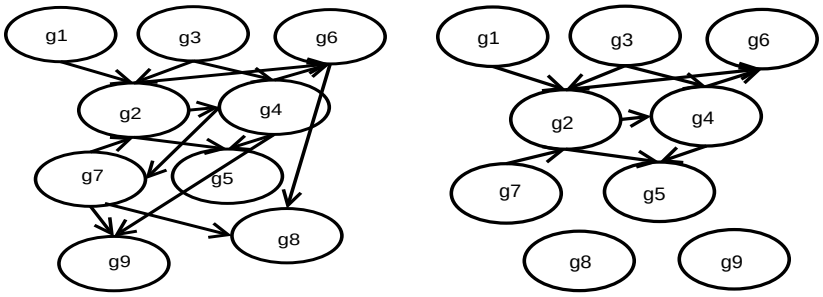
- Repeat the following procedure $B(= 1000)$ times (with index $b = 1, \ldots, B$):
  - Generate resampled (with replacement) versions of the times series X^N, Y^N, denoted by X_b^N, Y_b^N respectively.
  - Compute the statistic $\theta^b = \hat{I}(X_b^N \rightarrow Y_b^N)$.
- Construct an empirical CDF (cumulative distribution function) from these bootstrapped sample statistics, as $F_\Theta(\theta) = P(\Theta \leq \theta) = \frac{1}{B} \sum_{b=1}^{B} I_{x \geq 0}(x = \theta - \theta^b)$, where I is an indicator random variable on its argument x.
- Compute the true detection statistic (on the original time series) $\theta_0 = \hat{I}(X^N \rightarrow Y^N)$ and its corresponding p-value ($p_0 = 1 - F_\Theta(\theta_0)$) under the empirical null distribution $F_\Theta(\theta)$.
- If $F_\Theta(\theta_0) \geq (1 - \alpha)$, then we have that the true DTI value is significant at level α, leading to rejection of null-hypothesis (no directional association).

We now demonstrate some results using the above developed methods for the unsupervised and supervised network inference problems for a synthetic dynamical system as well as for a true biological problem, below.

4. RESULTS ON SYNTHETIC NETWORK

4.1 Synthetic Network

A synthetic network is constructed in the following fashion: We assume that there are variables g_1, g_3 and g_7 (all of which are modeled as uniform random variables) which drive the remaining variables of a nine variable network. The evolution equations are as below. The noise term, ϵ_t, is chosen to have a gaussian distribution $\mathcal{N}(0, \sigma^2)$, with a standard deviation concordant with experimental variation.



(With DTI) **(with CoD)**

Figure 1. The synthetic network as recovered by (a) DTI and (b) CoD.

$$
\begin{aligned}
g_{2,t} &= \frac{1}{2}g_{1,t-1} + \frac{1}{3}g_{3,t-2} + g_{7,t-1} + \epsilon_t; \\
g_{4,t} &= g_{2,t-1}^2 + g_{3,t-1}^{1/2} + \epsilon_t; \\
g_{5,t} &= g_{2,t-2} + g_{4,t-1} + \epsilon_t; \\
g_{6,t} &= g_{4,t-1} + g_{2,t-2}^{1/2} + \epsilon_t; \\
g_{7,t} &= \frac{1}{2}g_{4,t-1}^{1/3} + \epsilon_t; \\
g_{8,t} &= \frac{1}{2}g_{6,t-1}^{1/3} + \frac{1}{3}g_{7,t-1}^{1/2} + \epsilon_t; \\
g_{9,t} &= \frac{2}{3}g_{4,t-1}^{2/3} + \frac{1}{4}g_{7,t-2}^{1/2} + \epsilon_t;
\end{aligned}
$$

For the purpose of comparison, we study the performance of the Coefficient of Determination (CoD) approach for directed influence network determination. The CoD allows the determination of association between two variables via a R^2 goodness-of-fit statistic. The methods of (Hashimoto[21] , Li[27]) are implemented on the time series data. Such a study would be useful to determine the relative merits of each approach. We believe that no one procedure can work for every application and the choice of an appropriate method would be governed by the application (here, the biological question) under investigation. Each of these methods use some underlying assumptions and if these are consistent with the question that we ask, then that method has utility.

As can be seen (Fig. 1), though CoD can detect linear lag influences, the strongly non-linear ones are missed. DTI detects these influences and exactly reproduces the synthetic network. Given the non-linear nature of transcriptional kinetics, this is essential for reliable network inference. DTI is also able to resolve loops and cycles ($g_3, [g_2, g_4], g_5$ and g_2, g_4, g_7, g_2). Based on these observations, we examine the networks inferred using DTI in both the supervised and unsupervised settings.

4.2 Supervised Network Inference

Fig. 2 presents the results of upstream supervised network inference using DTI. Fig. 2(a) represents the graph upstream and downstream of g_7 based on the evolution equations above. Fig. 2(b) presents the variables that are found to be upstream of g_7 using DTI from the multivariate replicate sample trajectories. As can be observed, DTI correctly finds the the upstream effectors. Additionally, the DTI value can discriminate the strength of the various effectors. As shown in Fig. 2(b), the rank ordering of the DTI is $g_4 > g_3 > g_2 > g_1$ which completely correlates with the separation of the corresponding variable from g_7. Thus, DTI enables us to query both strength and significance for *any* pairwise variable relationship as opposed to those that are recovered from the data. As an example, the strength and significance of $I(g_4 \rightarrow g_7)$ is shown in Fig. 3.

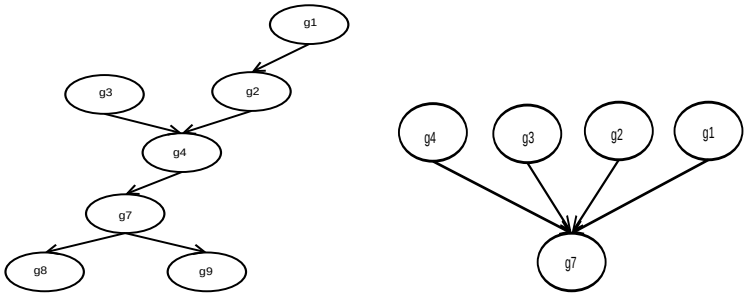


Figure 2. (a) The upstream and downstream effectors of g_7 from network dynamics (b) upstream effectors of g_7 from DTI.

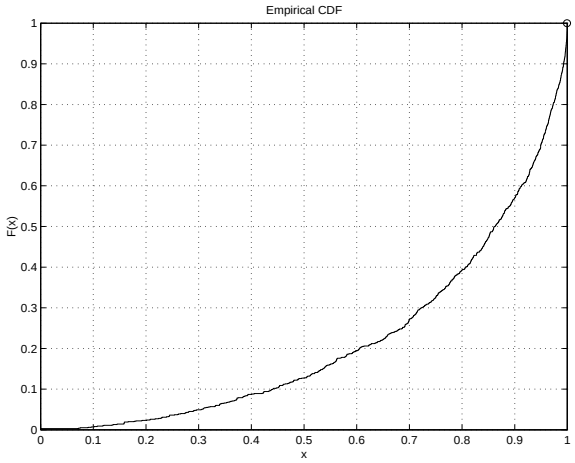


Figure 3. Cumulative Distribution Function for bootstrapped $I(g_4 \rightarrow g_7)$ over all permutations of the time series data. The true value of $I(g_4 \rightarrow g_7) = 0.9991$.

5. NETWORK DISCOVERY FOR BIOLOGICAL APPLICATIONS

The primary motivation for this work has been the need to infer directed dependencies between genes based on their expression data from microarray experiments (Beal[5] ,Rangel[38]). Microarrays are chips that can be used to simultaneously profile the expression of all genes in a cell. Most typically, they profile gene expression (via mRNA abundance) under various stimuli over time – thereby yielding a time series of expression for each gene. Using such data, biologists are deeply interested to find inter-dependencies among the genes so as to generate experimentally testable hypotheses about underlying biological processes. Since one gene might have role in influencing the behavior of another gene, thereby establishing a directed dependence, the inference of such directed networks has far reaching implications. In these examples, we will explain the use of DTI to infer such dependencies in both supervised and unsupervised settings. Further details are given under each head.

5.1 Directed Network Inference: *Gata3* regulation in early kidney development

Biologists have an interest in influence networks that might be active during organ development. Advances in laser capture microdissection coupled with those in microarray methodology have enabled the investigation of temporal profiles of genes putatively involved in these embryonic processes. Forty seven genes are expressed differentially between the ureteric bud and metanephric mesenchyme (Potter[46]) and putatively involved in bud branching during kidney development. The expression data (Grimmond[9]) temporally profiles kidney development from day 10.5 dpc to the neonate stage. The influence network amongst these genes is shown below (Fig. 4). Several of the presented interactions are biologically validated and there is an interest to confirm the novel ones pointed out in the network. The annotations of some of these genes are given below (Table. 1).

Some of the interactions that have been experimentally validated include the *Rara-Mapk1*[3] , *Pax2-Gata3*[18] and *Agtr-Pax2*[52] interactions. We note that this result clarifies the application of DTI for network inference

in an unsupervised manner - i.e. discovering interactions revealed by data rather than examining the strengths of interactions known a priori. Such a scenario will be explored later (Sec: 5.3). We note that though several interaction networks are recovered, we only show the largest network including *Gata3*, because this is the gene of interest in this study.

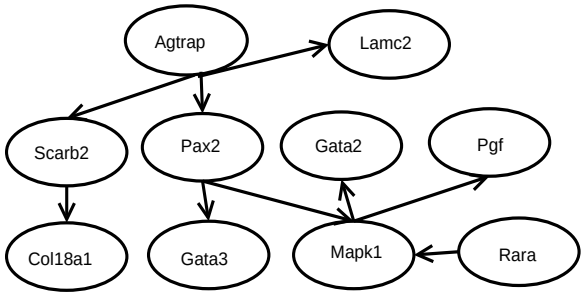


Figure 4. Overall Influence network using DTI during early kidney development.

Table 1. Functional annotations (*Entrez Gene*) of some of the genes co-expressed with *Gata2* and *Gata3* during nephrogenesis.

Gene Symbol	Gene Name	Possible Role in Nephrogenesis (Function)
Rara	Retinoic Acid Receptor	crucial in early kidney development
Gata2	*GATA* binding protein 2	several aspects of urogenital development
Gata3	*GATA* binding protein 3	several aspects of urogenital development
Pax2	Paired Homeobox-2	conversion of MM precursor cells to tubular epithelium
Lamc2	Laminin	Cell adhesion molecule
Pgf	Placental Growth Factor	Arteriogenesis, Growth factor activity during development
Col18a1	collagen, type $XVIII$, alpha 1	extracellular matrix structural constituent, cell adhesion
Agtrap	Angiotensin II receptor-associated protein	Ureteric bud cell branching

5.2 Directed Network Inference: T-cell Activation

To clarify the validity of the presented approach, we present a similar analysis on another data set - the T-cell expression data,[38] in Fig. 5. This data represents the expression of various genes after T-cell activation using stimulation with phorbolester PMA and ionomycin. The dataset contains the profiles of about 58 genes over 10 time points with 44 replicate measurements for each time point.

Several of these interactions are confirmed in earlier studies (Rangel[38] , Ezzat[16] , Zhang[53] , Rogoff[40]) and again point to the strength of DTI in recovering known interactions. The annotation of some of these genes are given in Table. 2. We note that the network of Fig. 5 shows the largest influence network (containing *Gata3*) that can be recovered. *Gata3* is involved in T-cell development as well as kidney development and hence it is interesting to see networks relevant to each context in Figs. 4 and 5. Also, these 58 genes relevant to T-cell activation are very different from those for kidney development, with fairly low overlap. For example this list does not include *Pax2* (which is relevant in the kidney development data).

5.3 Supervised Network Discovery

Based on *apriori* biological knowledge (such as literature) several molecules can be implicated in the regulation of a target gene (Kreiman[24] , Fraenkel[28]). Biologists are interested to explore this restricted space of possible effectors for possible influences on the target gene. An influence metric like DTI, if useful, would be invaluable to query any gene-gene relationship based on expression data.

For the kidney development case, we study the *Pax2-Gata3* interaction, and show the cumulative distribution function of the bootstrapped detection statistic (Fig. 3) as well as the position of the true DTI estimate in relation to the overall histogram. With the obtained density estimate of the *Pax2-Gata3* interaction, shown in Fig. 6, we can find significance values of the true DTI estimate in relation to the bootstrapped null distribution.

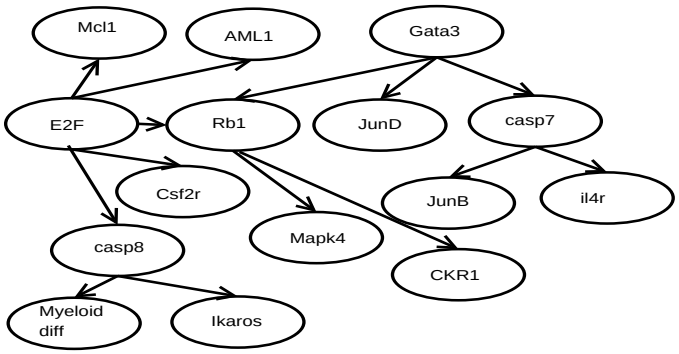


Figure 5. DTI based T-cell network.

Table 2. Functional annotations of some of the genes following T-cell activation.

Gene Symbol	Gene Name	Possible Role in T-cell activation (Function)
Casp7	Caspase 7	Involved in apoptosis
JunD	Jun D proto-oncogene	regulatory role of in T lymphocyte proliferation and Th cell differentiation
CKR1	Chemokine Receptor 1	negative regulator of the antiviral CD8+ T cell response
Il4r	Interleukin 4 receptor	inhibits *IL4*-mediated cell proliferation
Mapk4	Mitogen activated kinase 4	Signal transduction
AML1	acute myeloid leukemia 1; aml1 oncogene	CD4 silencing during T-cell differentiation
Rb1	Retinoblastoma 1	Cell cycle control

An experimental validation of this is presented in (Dressler[14] , Bouchard[18]), thereby indicating that DTI produces results that are consistent with experimental results.

6. SUMMARY OF ALGORITHM

We now present two versions of the DTI algorithm, one which involves an inference of general influence network between all variables of interest (*unsupervised-DTI*) and another, a focused search for effector variables which influence one particular variable of interest (*supervised-DTI*).

Our proposed approach using (*supervised-DTI*) for determining the effectors for variable B is as follows:

- Identify the G variables $(A_1, A_2, \ldots, A_G)$, based on required phenotypical characteristic using fold change studies. Preprocess the variable expression profiles by normalization and interpolation, if necessary. Assuming that there are N points for each variable, entropy estimation is used to compute the terms in the DTI expression (Eqn. 2).
- For each pair of variables A_i and B among these G variables :
  {
  - Find $DTI(A_i, B) = I(A_i^N \rightarrow B^N)$, and the normalized DTI from A_i to B, $\rho_{DTI}(A_i, B) = \sqrt{1 - e^{-2I(A_i^N \rightarrow B^N)}}$.
  - Bootstrap resampling over the data points of A_i and B yields a null distribution for $DTI(A_i, B)$. If the true $DTI(A_i, B)$ is greater than the 95% upper limit of the confidence interval (CI) from this null histogram, infer a potential influence from A_i to B.
  - The value of the normalized DTI from A_i to B gives the putative strength of interaction/influence.
  - Every variable A_i which is potentially influencing B is an 'effector'. This search is done for each variable A_i among these G variables $((A_1, A_2, \ldots, A_G))$.

  }

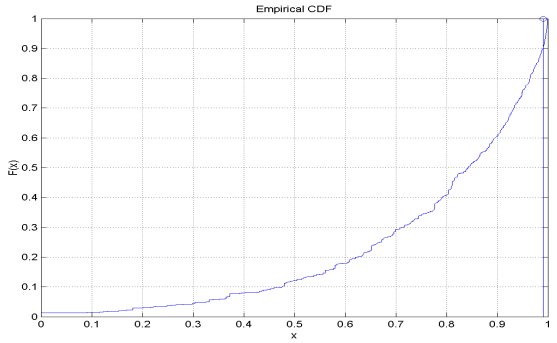


Figure 6. Cumulative Distribution Function for bootstrapped $I(Pax2 \rightarrow Gata3)$. The true value of $I(Pax2 \rightarrow Gata3) = 0.9911$.

Note: We note that, in *supervised-DTI*, the choice of potential effectors for a target variable (gene) is based on only those variables (TFs) that have a suspected dependence (such as a binding site at the target gene's promoter, Kreiman[24] , Fraenkel[28]). In this sense, *supervised-DTI* aims to reduce the overall search space based on biological prior knowledge.

For *unsupervised DTI*, we adapt the above approach for every pair of variables (A_i, B) in the list, noting that $DTI(A_i, B) \neq DTI(B, A_i)$. In this case we are not looking at any interaction in particular, but are interested in the entire influence network that can be potentially inferred from the given time series expression data. The network adjacency matrix has entries depending on the direction of influence and is related to the strength of influence as well as control of false discovery rate (FDR). The Benjamini-Hochberg procedure[6] is used to screen each of the $M(= G(G-1))$ hypotheses (both directions) during network discovery amongst G variables.

Briefly, the FDR procedure controls the expected proportion of false positives among the total number of rejections rather than just the chance of false positives[44] . It tolerates more false positives, and allows fewer false negatives.

- The p-values of the various edges $(1, 2, \ldots, M)$ are ranked from lowest to highest, all satisfying the original significance cut-off of $p = 0.05$. The ranked p-values are designated as $p_{(1)}, p_{(2)}, \ldots, p_{(M)}$.
- For $j = 1, 2, \ldots, M$, the null hypothesis (no edge) H_j is rejected at level α if $p_{(j)} \leq \frac{j}{M}\alpha$.
- All the edges with p-value $\leq p_{(j)}$ are retained in the final network.

In Table. 6, we compare the various contemporary methods of directed network inference. Recent literature has introduced several interesting approaches such as graphical gaussian models (GGMs), coefficient of determination (CoD), state space models (SSMs) for directed network inference. This comparison is based primarily on expectations from such inference procedures - that we would like any such metric/procedure to:

- Resolve cycles in recovered interactions.
- Be capable of resolving directional and potentially non-linear interactions. This is because interactions amongst genes involve non-linear kinetics.
- Be a non-parametric procedure to avoid distributional assumptions (noise etc).
- Be capable of recovering interactions that the application requires. Rather than use a method that discovers interactions underlying the data purely, the biologist should be able to use prior knowledge. For example, a biologist can examine the strength and significance of a known interaction and use this as a basis for finding other such interactions.

From the above comparisons, we see that DTI is one metric which can recover interactions under all these considerations.

Table 3. Comparison of various network inference methods.

Method	Resolve Cycles	Non -linear framework	Search for interaction	Non -parametric framework
SSM (Beal[5] , Rangel[38])	Y	Y	N	Y
CoD (Hashimoto[21])	N	N	Y	N
GGM (Strimmer[35])	N	Y	N	N
DTI (Rao[39])	Y	Y	Y	Y

7. CONCLUSIONS

In this work, we have proposed a principled methodology, using information theory for the discovery of influences among variables of a dynamical system based on replicated multivariate time series data. This has applications in biological networks, econometrics and neuroscience.

The proposed metric, directed time information (DTI) generalizes the partial correlation measure and can be used for influence discovery in both supervised and unsupervised settings. Using the biological networks as an example, we have shown the superiority of the DTI metric to other competitive metrics. We note that multiple realizations for the system trajectories are necessary for DTI to be viable for these applications. Also, DTI is able to resolve variable influences for very short time series data (such as those generated in microarray experiments) wherein the interval between sampling instants is much higher than in other applications (such as event-related time profiles in neuroscience).

Additionally, several other modifications can be introduced in this framework in the future. Since DTI computation is expensive (marginal and joint entropies have to be estimated for each n variable subset), we can use mutual information as a prior to reduce the search space to only those variables that have strong MI. Additionally, DTI can also be used to obtain edge priors for other network learning paradigms like bayesian networks (Woolf[51]). Another extension of this work is the comparison of DTI with other recent metrics like directed transinformation (Williams[43] ,[42]) or predictive information (Bialek[4]) on such evolving time series to understand the particular scenarios wherein one method might be more useful than another.

APPENDIX: A NORMALIZED DTI MEASURE

In this section, an expression for a 'normalized DTI coefficient' is derived. This is useful for a meaningful comparison across different criteria during network inference. The purpose of this section is to establish some connections between quantities like MI, DTI, and correlation. In this section, we use X, Y, Z for X^N, Y^N and Z^N interchangeably, i.e $X \equiv X^N$, $Y \equiv Y^N$, and $Z \equiv Z^N$.

By the definition of DTI, we can see that $0 \leq I(X^N \to Y^N) \leq I(X^N;Y^N) < \infty$. The normalized measure ρ_{DTI} should be able to map this large range ($[0,\infty]$) to $[0,1]$. We recall that the multivariate canonical correlation is given by Gubner[19] : $\rho_{X^N;Y^N} = \Sigma_{X^N}^{-1/2}\Sigma_{X^NY^N}\Sigma_{Y^N}^{-1/2}$ and this is normalized having eigenvalues between 0 and 1. We also recall that, under a Gaussian distribution on X^N and Y^N, the joint entropy $H(X^N;Y^N) = -\frac{1}{2}\ln(2\pi e)^{2N}|\Sigma_{X^NY^N}|$, where $|A|$ is the determinant of matrix A, $\Sigma_{X^NY^N}$ denotes the covariance matrix, computed as $\Sigma_{X^NY^N} = \frac{1}{R-1}X^TY$, indicating that there are R replicates of the X,Y time series, each of length N.

Thus, for $I(X^N;Y^N) = H(X^N)+H(Y^N)-H(X^N,Y^N)$, the expression for mutual information, under jointly Gaussian assumptions on X^N and Y^N, becomes, $I(X;Y) = -\frac{1}{2}\ln(\frac{|\Sigma_{X^NY^N}|^2}{|\Sigma_{X^N}|.|\Sigma_{Y^N}|}) = -\frac{1}{2}\ln(1-\rho^2_{X^N;Y^N})$. Hence, a straightforward transformation is normalized MI, $\rho_{MI} = \sqrt{1-e^{-2I(X^N;Y^N)}} = \sqrt{1-e^{-2\sum_{i=1}^{N} I(X^N;Y_i|Y^{i-1})}}$. A connection with Joe[23] , can thus be immediately seen.

With this, ρ_{MI} is normalized between $[0,1]$ and gives a better absolute definition of dependency that does not depend on the unnormalized MI. We will use this definition of normalized information coefficients in the present set of simulation studies.

For constructing a normalized version of the DTI, we can extend this approach, from Geweke[17] . Consider three random vectors $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, each of which are identically distributed as $\mathcal{N}(\mu_X, \Sigma_{XX})$, $\mathcal{N}(\mu_Y, \Sigma_{YY})$, and $\mathcal{N}(\mu_Z, \Sigma_{ZZ})$ respectively. We also have,

$$(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \sim \mathcal{N}\left[\begin{pmatrix} \mu_X \\ \mu_Y \\ \mu_Z \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} & \Sigma_{XZ} \\ \Sigma_{YX} & \Sigma_{YY} & \Sigma_{YZ} \\ \Sigma_{ZX} & \Sigma_{ZY} & \Sigma_{ZZ} \end{pmatrix}\right]$$

Their partial correlation $\delta_{YX|Z}$ is then given by, $\delta_{YX|Z} = \sqrt{\frac{a_2^2}{a_1 a_3}}$ with, $a_1 = \Sigma_{YY} - \Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZY}$, $a_2 = \Sigma_{YX} - \Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}$, $a_3 = \Sigma_{XX} - \Sigma_{XZ}\Sigma_{ZZ}^{-1}\Sigma_{ZX}$.

Recalling results from conditional Gaussian distributions, these can be denoted by: $a_1 = \Sigma_{Y|Z}, a_2 = \Sigma_{XY|Z}$ and $a_3 = \Sigma_{X|Z}$. Thus, $\delta_{YX|Z} = \Sigma_{Y|Z}^{-1/2}\Sigma_{XY|Z}\Sigma_{X|Z}^{-1/2}$. Extending the above result from the mutual information to the directed information case, we have, $\rho_{DTI} = \sqrt{1 - e^{-2\sum_{i=1}^{N} I(X^i;Y_i|Y^{i-1})}}$.

We recall the primary difference between MI and DTI, (note the superscript on X):
MI: $I(X^N;Y^N) = \sum_{i=1}^{N} I(X^N;Y_i|Y^{i-1})$.
DTI: $I(X^N \to Y^N) = \sum_{i=1}^{N} I(X^i;Y_i|Y^{i-1})$.

Having found the normalized DTI, we ask if the obtained DTI estimate is significant with respect to a 'null DTI distribution' obtained by random chance. This is addressed in Section 3.

We note that though the normality assumption was used to show the connection between information and correlation, this distributional assumption is not used anywhere in the original DTI metric formulation and computation during its application to network inference.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the NIH under award 5R01-GM028896-21 (J.D.E). We would like to thank Prof. Sandeep Pradhan and Mr. Ramji Venkataramanan for useful discussions on Directed Information. We are very grateful to Prof. Erik Learned-Miller for sharing his code for higher-order entropy estimation, and to Prof. Bruce Aronow and Prof. Sean Grimmond for the kidney expression data.

REFERENCES

[1] Aerts S, Thijs G, Coessens B, Staes M, Moreau Y, De Moor B., "Toucan: deciphering the cis-regulatory logic of coregulated genes", *Nucleic Acids Res.* 2003 Mar 15;31(6):1753-64.

[2] Moonen C.T.W. (Ed), Bandettini P.A. (Ed), Functional MRI (Medical Radiology / Diagnostic Imaging), Springer, 2000.

[3] Balmer JE, Blomhoff R., "Gene expression regulation by retinoic acid", *J. Lipid Res.* 2002 Nov;43:11:1773-808.

[4] Bialek W, Nemenman I, Tishby N, " Predictability, complexity, and learning.", *Neural Comput.* 2001 Nov ;13 (11):2409-63

[5] Beal MJ, Falciani F, Ghahramani Z, Rangel C, Wild DL.,"A Bayesian approach to reconstructing genetic regulatory networks with hidden factors", *Bioinformatics.* 2005 Feb 1;21(3):349-56.

[6] Benjamini, Y. and Hochberg, Y., "Controlling the false discovery rate: A practical and powerful approach to multiple testing", *J. Roy. Statist. Soc. Ser. B.* 1995; 57:289-300.

[7] Brophy PD, Ostrom L, Lang KM, Dressler GR., "Regulation of ureteric bud outgrowth by Pax2-dependent activation of the glial derived neurotrophic factor gene", *Development.* 2001 Dec;128(23):4747-56.

[8] Challen GA, Martinez G, Davis MJ, Taylor DF, Crowe M, Teasdale RD, Grimmond SM, Little MH., "Identifying the molecular phenotype of renal progenitor cells.", *J Am Soc Nephrol.* 2004 Sep;15(9):2344-57.

[9] Challen G, Gardiner B, Caruana G, Kostoulias X, Martinez G, Crowe M, Taylor DF, Bertram J, Little M, Grimmond SM., "Temporal and spatial transcriptional programs in murine kidney development", *Physiol Genomics.* 2005 Oct 17;23(2):159-71.

[10] Clarke JC, Patel SR, Raymond RM, Andrew S, Robinson BG, Dressler GR, Brophy PD., "Regulation of c-Ret in the developing kidney is responsive to Pax2 gene dosage.",*Hum Mol Genet.* 2006 Dec 1;15(23):3420-8.

[11] Cohen HT, Bossone SA, Zhu G, McDonald GA, Sukhatme VP.,"Sp1 is a critical regulator of the Wilms' tumor-1 gene", *J Biol Chem.* 1997 Jan 31;272(5):2901-13.

[12] Cover T.M, Thomas J.A, "Elements of Information Theory", Wiley-Interscience, 1991.

[13] Davies J, "Intracellular and extracellular regulation of ureteric bud morphogenesis", *Journal of Anatomy* 198 (3), 257264. (2001)

[14] Dressler, G.R. and Douglas, E.C. , "Pax-2 is a DNA-binding protein expressed in embryonic kidney and Wilms tumor", *Proc. Natl. Acad. Sci. USA* 89: 1179-1183, 1992.

[15] Effron B, Tibshirani R.J, An Introduction to the Bootstrap (Monographs on Statistics and Applied Probability), Chapman & Hall/CRC, 1994.

[16] Ezzat S, Mader R, Yu S, Ning T, Poussier P, Asa SL., "Ikaros integrates endocrine and immune system development", *J Clin Invest.* 2005 Apr;115(4):844-8.

[17] Geweke J., "The Measurement of Linear Dependence and Feedback Between Multiple Time Series," *Journal of the American Statistical Association*, 1982, 77, 304-324. (With comments by E. Parzen, D. A. Pierce, W. Wei, and A. Zellner, and rejoinder)

[18] Grote D, Souabni A, Busslinger M, Bouchard M., "Pax 2/8-regulated Gata3 expression is necessary for morphogenesis and guidance of the nephric duct in the developing kidney"., *Development.* 2006 Jan;133(1):53-61.

[19] Gubner J. A., Probability and Random Processes for Electrical and Computer Engineers, Cambridge, 2006.

[20] Smith, V., Yu, J., Smulders, T., Hartemink, A., Jarvis, E. (2006) Computational Inference of Neural Information Flow Networks. PLoS Computational Biology, 2, November 2006, pp. 14361449.

[21] Hashimoto RF, Kim S, Shmulevich I, Zhang W, Bittner ML, Dougherty ER., "Growing genetic regulatory networks from seed genes",*Bioinformatics.* 2004 May 22;20(8):1241-7.

[22] Hastie T, Tibshirani R, The Elements of Statistical Learning, Springer 2002.

[23] H. Joe., "Relative entropy measures of multivariate dependence", *J. Am. Statist. Assoc.*, 84:157164, 1989.

[24] Kreiman G., "Identification of sparsely distributed clusters of cis-regulatory elements in sets of co-expressed genes"., *Nucleic Acids Res.* 2004 May 20;32(9):2889-900.

[25] Miller E., "A new class of entropy estimators for multi-dimensional densities", *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2003.

[26] Learned-Miller E., "Hyperspacings and the estimation of information theoretic quantities", UMass Amherst Technical Report 04-104, 2004.

[27] Li H, Sun Y, Zhan M.,"Analysis of Gene Coexpression by B-Spline Based CoD Estimation",*EURASIP J Bioinform Syst Biol.* 2007;:49478.

[28] MacIsaac KD, Fraenkel E., "Practical strategies for discovering regulatory DNA sequence motifs".,*PLoS Comput Biol.* 2006 Apr;2(4):e36.

[29] Majumdar A, Vainio S, Kispert A, McMahon J, McMahon AP.,"Wnt11 and Ret/Gdnf pathways cooperate in regulating ureteric branching during metanephric kidney development", *Development.* 2003 Jul;130(14):3175-85.

[30] Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A., "ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context", *BMC Bioinformatics.* 2006 Mar 20;7 Suppl 1:S7.

[31] H. Marko, "The Bidirectional Communication Theory - A Generalization of Information Theory", *IEEE Transactions on Communications*, Vol. COM-21, pp. 1345-1351, 1973.

[32] J. Massey, "Causality, feedback and directed information," in *Proc. 1990 Symp. Information Theory and Its Applications (ISITA-90)*, Waikiki, HI, Nov. 1990, pp. 303305.

[33] Nemenman, F Shafee, and W Bialek. "Entropy and inference, revisited.", In TG Dietterich, S Becker, and Z Ghahramani, editors, *Advances in Neural Information Processing Systems* 14, Cambridge, MA, 2002. MIT Press.

[34] Nemenman I,. "Information theory, multivariate dependence, and genetic network inference.", Technical Report NSF-KITP-04-54, KITP, UCSB, 2004.

[35] Opgen-Rhein, R., and Strimmer K., "Using regularized dynamic correlation to infer gene dependency networks from time-series microarray data", *Proc. of Fourth International Workshop on Computational Systems Biology, WCSB*, 2006.

[36] Paninski, L. , "Estimation of entropy and mutual information", *Neural Computation* 15: 1191-1254, 2003.

[37] J. Ramsay, B. W. Silverman, Functional Data Analysis (Springer Series in Statistics), Springer 1997.

[38] Rangel C, Angus J, Ghahramani Z, Lioumi M, Sotheran E, Gaiba A, Wild DL, Falciani F, "Modeling T-cell activation using gene expression profiling and state-space models", *Bioinformatics*, 20(9),1361-72, June 2004.

[39] Rao A, Hero AO, States DJ, Engel JD, "Inference of biologically relevant Gene Influence Networks using the Directed Information Criterion", *Proc. of the IEEE Conference on Acoustics, Speech and Signal Processing*, 2006.

[40] Rogoff HA, Pickering MT, Frame FM, Debatis ME, Sanchez Y, Jones S, Kowalik TF., "Apoptosis associated with deregulated E2F activity is dependent on E2F1 and Atm/Nbs1/Chk2", *Mol Cell Biol.* 2004 Apr;24(7):2968-77.

[41] Ryan G, Steele-Perkins V, Morris JF, Rauscher FJ, Dressler GR.,"Repression of Pax-2 by WT1 during normal kidney development", *Development.* 1995 Mar;121(3):867-75.

[42] H. Shevrin, W.J. William, R.E. Marshall, R.K. Hertel, J.A. Bond, and L.A. Brakel, Event-Related Potential Indicators of the Dynamic Unconscious, Consciousness and Cognition, Vol. 1, pp. 340-366, 1992.

[43] W.J. Williams, H. Shevrin and R.E. Marshall, Information Modeling and Analysis of Event Related Potentials, IEEE Transactions on Biomedical Engr., Vol. BME-34, No. 12, pp. 928-937, Dec. 1987.

[44] Schfer, J., and K. Strimmer, "An empirical Bayes approach to inferring large-scale gene association networks", *Bioinformatics*,Oct 2004.

[45] Schneidman E, Still S, Berry MJ II, and Bialek W,"Network information and connected correlations", *Phys. Rev. Lett.*, 91, p. 238701, (2003)

[46] Schwab K, Patterson LT, Aronow BJ, Luckas R, Liang HC, Potter SS., "A catalogue of gene expression in the developing kidney", *Kidney Int.* 2003 Nov; 64(5):1588-604.

[47] Stuart RO, Bush KT, Nigam SK, "Changes in gene expression patterns in the ureteric bud and metanephric mesenchyme in models of kidney development", *Kidney International*,64(6),1997-2008,December 2003.

[48] Taraviras S, Monaghan AP, Schtz G, Kelsey G.,"Characterization of the mouse HNF-4 gene and its expression during mouse embryogenesis", *Mech Dev.* 1994 Nov;48(2):67-79.

[49] Venkataramanan R. , Pradhan S. S., "Source Coding With Feed-Forward: Rate-Distortion Theorems and Error Exponents for a General Source," *IEEE Transactions on Information Theory* , vol.53, no.6, pp.2154-2179, Jun. 2007.

[50] Willett R, Nowak R, "Complexity-Regularized Multiresolution Density Estimation", *Proc. Intl Symp. on Information Theory*, ISIT 2004.

[51] Woolf PJ, Prudhomme W, Daheron L, Daley GQ, Lauffenburger DA., "Bayesian analysis of signaling networks governing embryonic stem cell fate decisions", *Bioinformatics.* 2005 Mar;21(6):741-53.

[52] Zhang SL, Moini B, Ingelfinger JR., "Angiotensin II increases Pax-2 expression in fetal kidney cells via the AT2 receptor", *J Am Soc Nephrol.* 2004 Jun;15(6):1452-65.

[53] Zhang, DH, Yang L, and Ray A., "Differential responsiveness of the IL-5 and IL-4 genes to transcription factor GATA-3", *J Immunol* 161: 3817-3821, 1998.