Abstract

In this study we reconstruct predator-prey relationships from biomass time series of a simulated system of interacting species. To overcome the shortcomings of a static food webs representation we introduce a new model which accounts for both population and interaction dynamics. It is a derived version of the light-cone model from special relativity theory. To identify the existence of predator–prey relationships in the system we quantify the notion of distance in a food web. We use known measures from information theory, namely mutual information and transfer entropy, and we introduce a new measure based on causal states of point and patch predictors. To evaluate our results we compare the distances measured with a minimum distance measure from the underlying food web, and examine the accuracy of the measures in inferring the existence of the actual predator–prey relationships. First results show that our new measure based on causal states of point and patch predictors together with the transfer entropy measure outperform the mutual information measure in terms of distance accuracy. A threshold based method to estimate adjacent links shows similar results.

Computational mechanics and information measures in food webs

O. Bochmann[*], J. T. Lizier[†], J. Mahoney[‡], G. Obernosterer[§], J. Pahle[¶]

August 15, 2007

Contents

[*]Department of Ecology, Evolution and Organismal Biology, Iowa State University, 50011 Ames IA, USA; olaf@iastate.edu

[†]CSIRO Information and Communications Technology Centre, Locked Bag 17, North Ryde, NSW 1670, Australia; School of Information Technologies, The University of Sydney, NSW 2006, Australia; jlizier@it.usyd.edu.au

[‡]Physics Department, University of California, Davis, CA 95616, USA; jrmahoney@ucdavis.edu

[§]Institute of Molecular Biotechnology at the Austrian Academy of Sciences, Dr. Bohr-Gasse 3, 1030 Vienna, Austria; gregor@tbi.univie.ac.at

[¶]Bioinformatics and Computational Biochemistry, EML Research, Schloss-Wolfsbrunnenweg 33, 69118 Heidelberg, Germany; juergen.pahle@eml-r.villa-bosch.de

1 Introduction

In this study we are interested in reconstructing dynamic food webs from biomass time-series data of individual species in an ecosystem. The question is how relationships between species can be identified solely by knowing their population dynamics. We know that information theory provides a number of tools to identify these relationships as static measures (e.g. mutual information between species). However, we know that food webs are not static — the relationships between species change over time. Local information theoretic measures can reveal more about the dynamics in time than their averaged counterparts (e.g. see Lizier et al. [2007]), however they do not reveal the internal states of the system. A more ambitious goal is to capture the internal states of the species, the *structure* in the dynamics of these relationships, and represent these in an appropriate model which can also be used for *prediction* of the dynamics.

A food web is a set of related food chains in a particular ecosystem. A food chain is a set of related species. The relationship among the species is in such a way that each species feeds upon on the one below it and is eaten by the species above it in the chain. This is what is known as predator–prey relationship. The food web can be represented as a graph $G = (U, A)$ with node set $U = \{u_1, \ldots, u_n\}$ and arc set $A = \{a_{ij} : \forall i, j\}$. Each arc $a_{ij} \in A$ defines a predator–prey relationship from species u_i to species u_j. To represent cannibalism a special case for $i = j$ is included.

Each node in the food web represents a species and is associated to a quantity e.g. biomass. Arcs represent relationships between species and connect to a quantity e.g. biomass flow. These quantities are dynamic: they change over time. Also, they can become zero and extinguished species and predator–prey relationships disappear from the graph. Alternatively, new species may appear over time. This shows that the graph G is not a good way to visualize the dynamics of a food web. It is rather a static snapshot of the process or an aggregation over a certain period of time. We have also seen an animated version of the graph [Williams et al., 2007], where nodes and arcs shrink and expand to represent the value of there quantities. Obviously, an animation is difficult to illustrate on paper. Therefore we propose a different model, which is derived from special relativity theory.

To model the dynamics of the food web we adapt the light cone model for our particular network problem. In addition to the vertices and edges

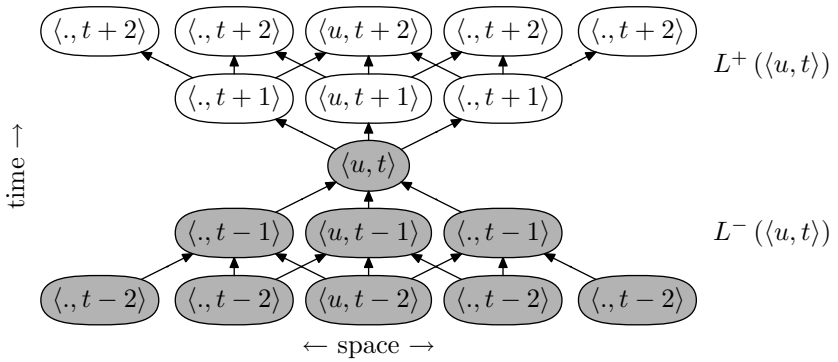

Figure 1: Light-cone diagram of an event at the point $\langle u, t\rangle$. Time runs vertically upwards. The shaded nodes denote the past light-cone of $\langle u, t\rangle$, the white ones the future light-cone. Space, here drawn as one-dimensional lattice, is an ordered set of $\mathcal{U}_i$, with $\mathcal{U}_i \subset \mathcal{U}$ and of $\mathcal{E}_j$, with $\mathcal{E}_j \subset \mathcal{E}$. [Bochmann, 2005]

of the graph G a time coordinate $t \in \mathbb{N}$ or $t \in \mathbb{Z}$ is introduced, creating a spatiotemporal lattice structure. Each combination of vertex and time is called a *point* $\langle u, t\rangle$. At each point, there is a random variable $X(\langle u, t\rangle)$ taking values in an alphabet set $\mathcal{A}$. The alphabet is a set of discrete symbols a e.g. the boxed biomass of the species. The signal propagation speed c in the network is assumed to be one hub per time unit (where a one hub signal propagation at a given time step is only possible between causal information sources and destinations, i.e. vertices connected by an edge in the graph). Each event has a space and a time coordinate just as the random variable $X(\langle u, t\rangle)$, where the space is defined by a particular species in graph G. As in figure 1, each event has a (virtual) double–cone structure attached to it, where the root vertex corresponds to the event itself. A surface crossing that vertex horizontally represents the present, and the position of the vertex is the position of the observer. By convention, "time" runs vertically in this diagram. The upward-directed cone opens to enclose the directions pointing towards events in the future (*future light-cone*). The downward-directed cone directions pointing towards events in the past (*past light-cone*).

Events in the past belong to the past light-cone of a point $\langle u, t\rangle$, if they could influence the field at that point. Its configuration is given by

$$L^{-}(\langle u, t\rangle) \equiv \bigcup_{\tau \geq 0} \left\{ X(\langle v, t-\tau\rangle), X(\langle (\mathrm{p}(v), v), t+\tau\rangle) \mid \bigvee_{k=0}^{k=c\tau} vE^k u \right\}. \quad (1)$$

Events which could influence the observer at any given time lie inside a concentric[1] ring on the past light-cone. Given the observer event is $\langle u, t\rangle$, all events $\langle v, \tau\rangle$ for $\tau < t$ and $\forall v : vE^k u = c(t - \tau)$ lie inside that ring. All other events in the past are too far in distance that the signal could reach the observer. For practical reasons, $\tau \leq \tau^-_{\max}$ limits how far the light-cones used for analysis extend into the past.

The future light-cone of a point $\langle u, t\rangle$ is the set of all points whose fields could be influenced by the field at $\langle u, t\rangle$ (excluding $\langle u, t\rangle$). Its configuration is given by

$$L^+ (\langle u, t\rangle) \equiv \bigcup_{\tau \geq 1} \left\{ X (\langle v, t + \tau\rangle) , X (\langle (\mathrm{p} (v) , v), t + \tau\rangle) \mid \bigvee_{k=0}^{k=c\tau} vE^k u \right\}. \quad (2)$$

Events which could be influenced by the observers state at a specific time in the future lie on a concentric ring on the future light-cone. All other future events are either too far in distance that the signal could reach them at this time or they are closer in distance and the signal would have reached them at some earlier time. The actions of two observers "meet" and may influence each other at a time, when their future light-cones overlap. For practical reasons, $\tau \leq \tau^+_{\max}$ limits how far light-cones extend into the future.

In attempting to recreate food web dynamics, we initially have no information on which species are directly connected to which other species. That is to say, we do not know which nodes are linked thereby allowing a one-hub signal propagation between them at any given time step. To perform any analysis, we have to assume that every species is directly linked to every other species. Examining a food web in the light cone model then means that the cone in fact becomes a concentric multi-dimensional ring only (there are no events inside the light cone since each node is linked to each other node at the next time step).

While it would be theoretically accurate, it becomes practically intractable to analyze or model the system as a whole or even individual nodes when one has to assume that every node is causally linked to every other node. Ideally, the alternative is to establish the links in the species network to allow a simplified but still accurate analysis for each species (note that the network could not be expected to form a regular lattice structure in general). In this paper we will describe analytic techniques which could be used to indicate the links in the species network. We will also describe a more simplified analytic technique to produce models of the dynamics at each species, which represent a compromise between accuracy and tractability.

[1] Constant signal propagation speed is assumed.

Reconstructing adjacent structure in a food web, that is to distinguish direct from indirect predator-prey relationships among the different species, is a nontrivial problem. We know from Wagner [2001] that the identification of adjacent interactions usually involves perturbation of the system. Since we are limited to observations only we will focus on reachability relationships rather then adjacent relationships.

2 Background

Population dynamics in general can be seen as a stochastic process, where one might be interested in the amount of information shared by the history and the future of the process. A food web represents the interaction of different species. To reconstruct the relationships between the species one might be interested in the information dynamics of their interactions. We will describe two information-theoretical measures which will be used to analyze the interactions. We are also interested in the structural pattern of the dynamics and of the interactions, in particular in developing models of the processes involved. Computational mechanics provides some promising approaches, and the models that it produces might provide a deeper understanding of dynamic food webs. We will outline the theory of causal states and ϵ-machine subsequently.

2.1 Mutual Information and Transfer Entropy

Mutual information by Shannon and Weaver [1964] is a measure of the amount of information that one random variable contains about another random variable. It is the average reduction in the uncertainty of one random variable X due to the knowledge of the other (fixing Y). The random variable Y carries information about X, if knowing Y makes X more certain on average than if Y were unknown. The mutual information $\mathrm{I}[X;Y]$ between two random variables X and Y is

$$\begin{aligned} \mathrm{I}[X;Y] &\equiv \mathrm{H}[X] - \mathrm{H}[X \mid Y] \\ &\equiv \mathrm{E}_{X,Y}\left[\log_2 \frac{\Pr(X=x, Y=y)}{\Pr(X=x)\Pr(Y=y)}\right]. \end{aligned} \tag{3}$$

It is the expected logarithm of the ratio between the actual joint distribution of X and Y, and the product of their marginal distributions.

If $\mathrm{H}[X] = 0$ or $\mathrm{H}[Y] = 0$, then there is no information to share and $\mathrm{I}[X;Y] = 0$. Independence of X and Y means there is no "communication"

between the variables and again $\mathrm{I}[X;Y] = 0$. Mutual information is symmetric in X and Y. Assuming the history light-cone $X = L^-(\langle u,t\rangle)$ and the future light-cone $Y = L^+(\langle u,t\rangle)$ as random variables, then the measure is the amount of information that the future process contains about the history. Alternatively, we can assume the history light-cone of one point $X = L^-(\langle u,t\rangle)$ and the history light-cone of another point $Y = L^-(\langle v,t\rangle)$ as random variables. Then the measure is the amount of information that one point contains about the other.

The mutual information has also been used as a measure of information transfer, however this is criticized by Schreiber Schreiber [2000] because mutual information is a static, asymmetric measure of shared information only. To address these inadequacies, Schreiber introduced the transfer entropy as a dynamic, asymmetric measure of information transfer Schreiber [2000]. The transfer entropy describes the amount of information that a source Y provides [2] about the next state of a destination X' that was not contained in the k past states of the destination $X^{(k)}$:

$$T_{Y\to X} = \mathrm{E}_{Y,X',X^{(k)}}\left[\log_2 \frac{\Pr\left(X' = x_{n+1}|X^{(k)} = x_n^{(k)}, Y = y_n\right)}{\Pr\left(X' = x_{n+1}|X^{(k)} = x_n^{(k)}\right)}\right]. \qquad (4)$$

The formulation is completely accurate in the limit $k \to \infty$, however since this is not practically computable, reasonable estimates can be made with finite values of k [Lizier et al., 2007]. The transfer entropy can also be represented as a conditional mutual information:

$$T_{Y\to X} = \mathrm{I}\left[Y; X' \mid X^{(k)}\right] = \mathrm{I}\left[X' \mid X^{(k)}\right] - \mathrm{I}\left[X' \mid Y, X^{(k)}\right]. \qquad (5)$$

Clearly, the transfer entropy quantifies the information added by the source about the next state of the destination in the context of the destination's past; as such, it is a dynamic, direction measure of information transfer. It is not explicitly a measure of causality; however where one has access to observational data only, perturbation-based techniques are not possible and transfer entropy represents a reasonable compromise to infer causality.

2.2 Causal States and ϵ-Machine

Computational mechanics is a way of measuring the complexity of an process. It provides a direct structural model of a system's intrinsic information

[2] We set the source length l (see Schreiber [2000]) to its default value of 1 so as to consider transfer from the immediately previous source value only.

processing. In other words, it explains how a system stores, transmits and manipulates information. To produce this model, we use a measuring device to observe the original process via a learning channel. The measurement projects the internal state of the process down to a lower dimensional discretized state. The problem for the observer is to infer the causal states and state transitions of the minimal stochastic model of the original process. This model is called ϵ-machine and provides insights on the intrinsic computation, the pattern and complexity of the original hidden process.

The fundamental idea about causal states is that it makes only sense to distinguish two different inputs if they actually create meaningfully different outputs. In the context of our initial problems this means: 1. We observe history sequences and try to predict future sequences. Two history sequences belong to the same causal state if we predict the same set of future sequences with the same probabilities for each. 2. We observe history sequences and are interested in predicting interactions between species in the web. Two history sequences belong to the same causal state, if they predict the same interaction.

Shalizi and Crutchfield [2001] developed the mathematical foundation of causal states. Formally, there is a distribution over future light-cones conditional to each past light-cone realization l^- at a point

$$\Pr\left(L^+\left(\langle u,t\rangle\right) \mid L^-\left(\langle u,t\rangle\right) = l^-\right) \equiv \Pr\left(L^+ \mid l^-\right). \tag{6}$$

To define the equivalence relationship, it is assumed that two pasts are equivalent if they have the same probability distribution of future light-cones.

$$l_1^- \sim l_2^- \Leftrightarrow \Pr\left(L^+ \mid l_1^-\right) = \Pr\left(L^+ \mid l_2^-\right). \tag{7}$$

The set of all pasts having the same distribution on the future light-cone is called an *equivalence class*. This allows one to define a local statistic $F = \eta\left(L^-\left(\langle u,t\rangle\right) = l^-\right)$, which is the equivalence class on the past light-cone. A good local statistic predicts something about what will happen in the future light-cone; here this is quantified as the mutual information between that local statistic and the future light-cone $\mathrm{I}\left[L^+; F\right]$. The data processing inequality bounds the information which can be conveyed in a statistic to

$$\mathrm{I}\left[l^+; F\right] \leq \mathrm{I}\left[l^+; l^-\right]. \tag{8}$$

A sufficient statistic F is one that reaches this boundary, and is thus as informative as the original data. The local causal state at a point $\langle u,t\rangle$ is the set of all past light-cones whose conditional distribution of future light-cones is the same and therefore it is the equivalence class at that point.

$$S\left(\langle u,t\rangle\right)=\epsilon\left(l^{-}\right)=\left\{\lambda:\Pr\left(l^{+}\mid\lambda\right)=\Pr\left(l^{+}\mid l^{-}\right)\right\} \tag{9}$$

The local causal states are *sufficient* statistic because $\Pr\left(l^{+}\mid\epsilon\left(l^{-}\right)\right)=\Pr\left(l^{+}\mid l^{-}\right)$, and so $\mathrm{I}\left[l^{+};\epsilon\left(l^{-}\right)\right]=\mathrm{I}\left[l^{+};l^{-}\right]$. They are predicting the same possible future with the same probabilities as the past light-cones would, which they contain. It is the coarsest set of predictively sufficient states. The past and the future light cone are *independent* given the local causal state.

Local causal states are *minimal* (optimality) because equivalent sufficient statistics $\eta\left(l_{1}^{-}\right)=\eta\left(l_{2}^{-}\right)$ with $\Pr\left(l^{+}\mid l_{1}^{-}\right)=\Pr\left(l^{+}\mid l_{2}^{-}\right)$ would belong to the same causal state $\epsilon\left(l^{-}\right)$. They would capture the minimal amount of information needed to predict the future of the process, which is how much information about the past of the process is relevant to predicting its future. Moreover, they are unique because any other state would just relabel the same state.

The causal states $\mathcal{S}=\left\{\mathcal{T}^{(s)},s\in\mathcal{A}\right\}$ together with the probability of transitions $T_{ij}^{(s)}$ between causal states are an ϵ-machine $\mathrm{M}=\{\mathcal{S},\mathcal{T}\}$, a minimal model capable of statistically reproducing the original process. The ϵ-machine is deterministic in the sense that given the symbol always leads to at most one single state.

With respect to our application to food webs, constructing ϵ-machines for single species would allow us to build models for the dynamics of each species. Ideally, this would be performed considering using light-cones containing all other species (since we do not know the relevant causal links), however as stated earlier this is computationally very difficult. Instead, we will construct ϵ-machines for single species using their single-dimensional time-series only, which models the dynamics observed when viewing each species in isolation.

Furthermore, we are also interested in an investigation of the *interaction* of different species in the food web. Shalizi [2003] has addressed the problem of predicting multiple vertices in the limit of the entire network, where only local causal states are available. Let us assume a connected set of points at a common time t (patch). As shown in figure 2, this patch has a past and a future light-cone (P^{-} and P^{+}), which are the unions of the cones of the constituent points. If the patch is defined as two spatially adjacent points $\langle u,t\rangle$ and $\langle v,t\rangle$, several regions in the joint light-cones can be defined. Regions that belong exclusively to point $\langle u,t\rangle$ are $U^{-}=L^{-}\left(\langle u,t\rangle\right)\setminus L^{-}\left(\langle v,t\rangle\right)$ and $U^{+}=L^{+}\left(\langle u,t\rangle\right)\setminus L^{+}\left(\langle v,t\rangle\right)$, overlapping regions that belong to point $\langle u,t\rangle$ and $\langle v,t\rangle$ are $C^{-}=L^{-}\left(\langle u,t\rangle\right)\cap L^{-}\left(\langle v,t\rangle\right)$

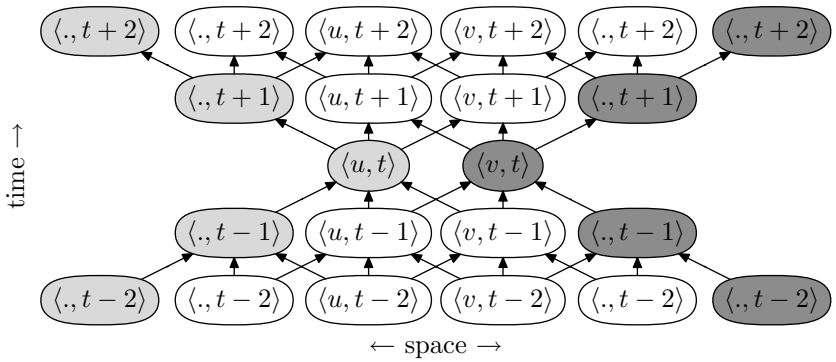

Figure 2: Two-point patch predictor: The light grey shaded points belong exclusively to the light-cones of point $\langle u, t\rangle$ (U^- and U^+), the dark gray shaded ones belong exclusively to the point $\langle v, t\rangle$ (V^- and V^+). The white points belong to the area of overlap (C^- and C^+). [Bochmann, 2005]

and $C^+ = L^+(\langle u, t\rangle) \cap L^+(\langle v, t\rangle)$ and regions that belong exclusively to point $\langle v, t\rangle$ are $V^- = L^-(\langle v, t\rangle) \setminus L^-(\langle u, t\rangle)$ and $V^+ = L^+(\langle v, t\rangle) \setminus L^+(\langle u, t\rangle)$.

When reconstructing a dynamic food web, we might come to the question of what kind of predator–prey relationship it is supposed to represent. Beside the static and aggregated versions we have discussed earlier, in a dynamic context one could be interested in how much a predator has influenced the prey population up to time t. This would be a quantity related to the region C^-. On the other hand, the region C^+ is related to the prediction problem: How much will a predator influence the current prey population in the future.

3 Data and Methods

The data for this study where generated using a simulation tool from Williams et al. [2007] with a food web called "Nic20withName". A static view of the food web is shown in figure 3, with predator–prey relationships marked by links in the diagram. It was initialized with 20 nodes and 41 interactions, including two self-interactions to model cannibalism. The underlying dynamic model was not known to us at the time of this analysis, with the motivation of the experiment being to attempt to model the dynamics using biomass data only.

The simulation was run for 117438 time steps. At each time step the biomass at each species and biomass-flow at each interaction where recorded.

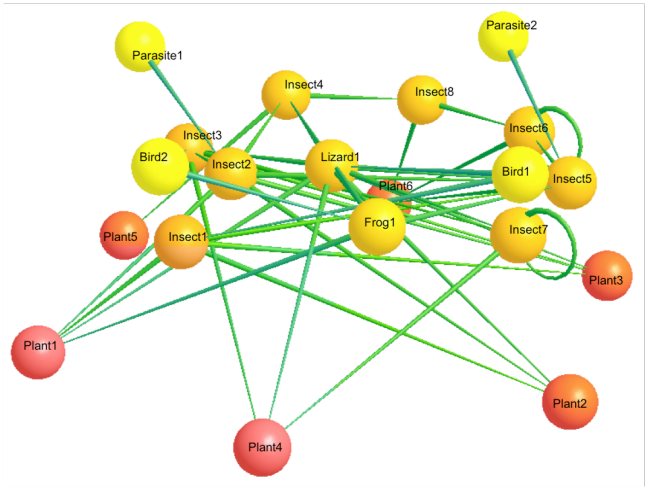

Figure 3: Example food web "Nic20withName" with 20 species and 41 predator-prey interactions. The names of the species are as informative as Plant1, Plant2, Plant3, Plant4, Plant5, Plant6, Insect1, Insect2, Insect3, Insect4, Insect5, Insect6, Insect7, Insect8, Bird1, Bird2, Frog1, Lizard1, Parasite1 and Parasite2. [Williams et al., 2007]

We had two extinction events shortly after the start of the simulation (Plant1, Plant5). To avoid any problems with non-stationarity we truncated the time series at this point. The corrected data-set consists of 18 biomass time series and 116823 time steps. To simulate the discretization of a measuring device we boxed the continuous data into a binary data-set at the median[3] of each time series ($0 \equiv low$, $1 \equiv high$). The 18 corrected and boxed time series were used during subsequent analysis. To simulate a realistic reconstruction scenario, the qualitative and quantitative interaction data (biomass-flow) are used only to evaluate our results.

Our experiments focus on using computational mechanics to discover topological patterns in the food web network, which is to our knowledge, a new approach. Ideally, we would reconstruct the ϵ-machine for the global light-cone configuration of a fully connected network, however computational resource and data constraints make this intractable. Intractable also is the creation of ϵ-machines for local light-cone configurations using full connectivity. Instead, we examine creating point predictor ϵ-machines for each node using their single-dimensional time-series, then investigate methods of inferring the food web network structure so as to determine the actual light-cone structure for each species and create more complete yet tractable ϵ-machines.

3.1 Point predictor ϵ-Machine reconstruction

As a first approximation, we construct an ϵ-machine for each species (a point predictor) using only the single-dimensional time-series biomass data for that species. For ϵ-machine reconstruction we used the algorithm from Shalizi and Shalizi [2003] with history length $m = 10$ and significance level $\alpha = 0.001$.

3.2 Inferring predator–prey relationships

In order to reconstruct more accurate ϵ-machine point predictor models for each species, we need to know which species should be included in the light-cone for the species under consideration. This requires knowledge of the links (i.e. predator–prey relationships) in the food web. To attempt to infer predator–prey relationships, construct three distance metrics to be measured between each species pair in the food web. A small distance could be used to infer the existence of a predator–prey relationship.

[3]Thanks to Tom Carter we know that this is the maximum entropy configuration.

3.2.1 Mutual Information distance measure

The mutual information was evaluated between the biomass data of each species pair, examining the single states of each species at common time steps. The mutual information represents a closeness measure between 0 and 1 bit; a distance metric is computed by subtracting this value from 1. Interestingly, a related study by Nichols [2005] used mutual information between two populations to study their relationship.

The use of mutual information as a distance measure is potentially problematic however because it represents statically shared information rather than necessarily an indication of a causal link in either direction between the species. Two species could have a high mutual information because of interactions both have with a shared third party rather than with each other.

3.2.2 Transfer entropy distance measure

An alternative distance measure can be obtained using the transfer entropy. It will not reflect shared information as the mutual information does, but a directional information transfer which is potentially more akin to effects in predator-prey relationships. The transfer entropy is measured between the biomass data of each species pair, examining the transfer from the previous value of the source variable to the next state of the destination, conditioned on the previous $k = 10$ states of the destination variable. The transfer entropy represents a directional closeness measure between 0 and 1 bit. A directional distance measure could be computed as for the mutual information by subtracting this value from 1. A symmetric distance measure on the other hand could be computed by taking the larger of the two transfer entropies and subtracting it from one; strength in either connection is enough to infer a predator–prey relationship.

3.2.3 ϵ-Machine patch predictor distance measure

Another measure of distance between two species is suggested here, based on statistical complexities of individual point predictors for pairs of nodes and the corresponding patch predictor for those same nodes.

We start with the assumption that observing the event at point $\langle u, t\rangle$ renders the point $\langle u, t+1\rangle$ in the future light-cone independent from the history light-cone. Next, having an ϵ-machine M_u for process u renders all points $\langle u, t+i\rangle$ in the future light-cone independent and determines point $\langle u, t+1\rangle$. We can do the same for a process v. Then, observing the event at point $\langle v, t\rangle$ and having the ϵ-machine M_v summarizes all information from

the history light-cone to predict $\langle v, t+i \rangle$ for $i = 2 \ldots \infty$[4] and determines point $\langle v, t+1 \rangle$. If we now assume that there is a region $C^{-*} \in C^{-}$ of the patch $\langle (u, v), t \rangle$ that contains structure which contributes to the prediction of $\langle u, t+i \rangle$ and $\langle v, t+i \rangle$, then this structure is part of both, M_u and M_v. We can now reconstruct an ϵ-machine $\mathrm{M}_{(u,v)}$ for processes u and v. Because of minimality of causal states, $\mathrm{M}_{(u,v)}$ contains structure coming from C^{-*} only one time, while M_u and M_v both carry this structure. This allows us to isolate the structure coming from C^{-*}, which can be quantified with statistical complexity C_μ as:

$$\mathrm{C}_\mu(u, v) \equiv \mathrm{C}_\mu(\mathrm{M}_u) + \mathrm{C}_\mu(\mathrm{M}_v) - \mathrm{C}_\mu(\mathrm{M}_{(u,v)}). \tag{10}$$

This new measure quantifies the history-structure common to both processes that is used to predict both of their futures independently. It is similar to mutual information in the sense that it is symmetric in u and v.

Again, for ϵ-machine reconstruction we used the algorithm from Shalizi and Shalizi [2003] with history length $m = 10$ and significance level $\alpha = 0.001$. We reconstructed patch predictors for each possible pair combination.

4 Results and discussion

4.1 Point predictor ϵ-machine reconstruction

We have reconstructed ϵ-machines for the single-dimensional time series of all species. One example for species "Insect6" is shown in figure 4. The nodes in the graph represent causal states and arcs represent state transitions. The emitted symbol a is labeled on each state transition together with the transition probability. Loops in the trajectory represent recursive patterns in the dynamic process. The size of the reconstructed patterns is bounded by the parameter m. We are using the inferred statistical complexity C_μ of each of this point predictors together with the patch predictor to create a distance measure.

To learn more about the dynamics of the process and the convergence of the reconstructed ϵ-machines to its original process we increased the considered history length m gradually and observed the model size. The diagram in figure 5 shows the number of causal states and figure 6 the statistical complexity C_μ as a function of history length m for all the species[5]. In case of

[4] note that this is only a subset of the future light-cone

[5] The labels in the diagram should read: Plant1, Plant2, Plant3, Plant4, Plant5, Plant6, Insect1, Insect2, Insect3, Insect4, Insect5, Insect6, Insect7, Insect8, Bird1, Bird2, Frog1,

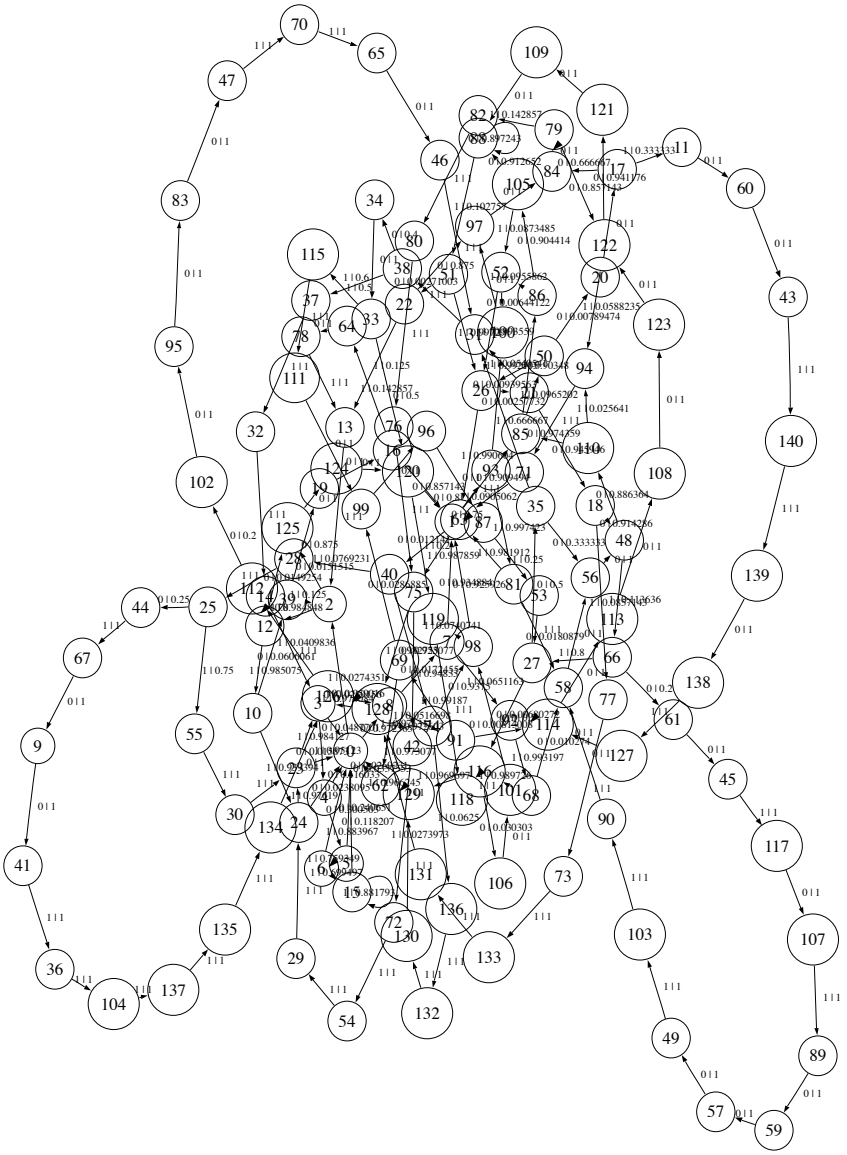

Figure 4: Example of one of the ϵ-machines reconstructed for the single-dimensional time series of the species "Insect6". CSSR inferred 141 states with $C_\mu = 4$ using a history of $m = 10$ at significance level $\alpha = 0.001$.

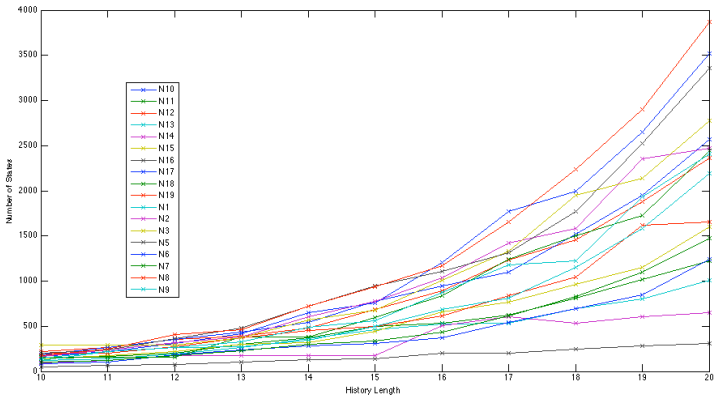

Figure 5: Number of inferred states as function of history length m for each of the species.

convergence we would expect some settling of size and complexity with increase of m. We can not see any sign of it in the range $10 < m < 20$. This could have several reasons. One reason could be that the processes have no finite model, e.g. chaotic process. Another reason could be that the data-sets are to short. CSSR tends to find non-existing patterns due to over-fitting.

4.2 Inferring predator–prey relationships

The results for the distance metrics computed between each pair of species are presented here in two-dimensional plots using the multi-dimensional scaling technique (MDS). This technique projects a set of n points, specified by their distance matrix, onto an m-dimensional space. For example, see a description of the MDS implementation for Matlab at Mathworks [2007].

These two dimensional projections of the computed distances between each node are shown for the mutual information, transfer entropy and ϵ-machine patch predictor measures respectively in Figures 7, 8 and 9.

To evaluate the relative success of the measures in inferring links in the food web, compare the distance projections to the actual links in the generating food web model in figure 3. These reconstructions appear to

Liyard1, Parasite1, Parasite2 instead of N0, N1, N2, N3, N4, N5, N6, N7, N8, N9, N10, N11, N12, N13, N14, N15, N16, N17, N18, N19.

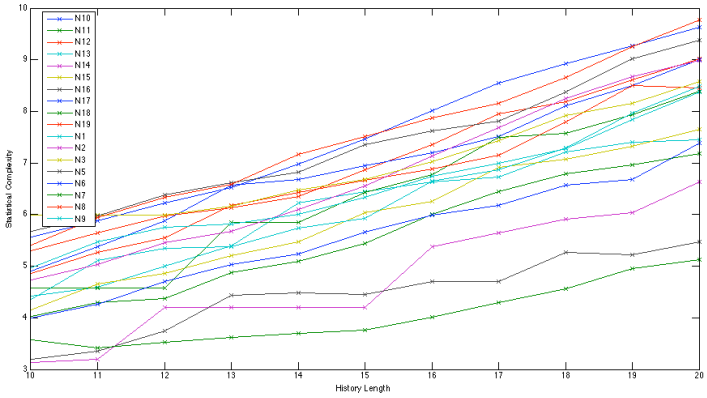

Figure 6: Statistical complexity C_μ as function of history length m for each of the species.

accurately represent some of the structure in the original food web, e.g. proximity of Parasite 1 and Plant 2 in figure 3 and figure 8, but miss others such as the Lizard 1 and Insect 4 relationship.

To *quantify* this evaluation, we compare the distances measured by each metric to some notion of actual distance in the original food web. The distance in a graph is frequently expressed as shortest path between two nodes: we take the shortest paths between each pair using the known predator–prey links in the original food web model to create a reference distance matrix here. An MDS plot generated from the shortest path distances can be seen in figure 10. This allows a visual comparison with the original food web in figure 3 and with our generated distance measures in figure 7, figure 8 and figure 9.

We then make a quantitative comparison by normalizing each distance matrix and subsequently taking the mean square error of the pair-wise distances computed by each of our distance metrics compared to the shortest path length in the known network structure. We found the normalized mean square errors were as follows (in square normalized units): mutual information measure: 0.2056; transfer entropy measure: 0.1481; patch predictor measure: 0.1514. This suggests that the patch predictor method and the transfer entropy method produce the more accurate reconstruction of the distances between each species then the mutual information method. These measures are not necessarily a correct representation of accuracy here how-

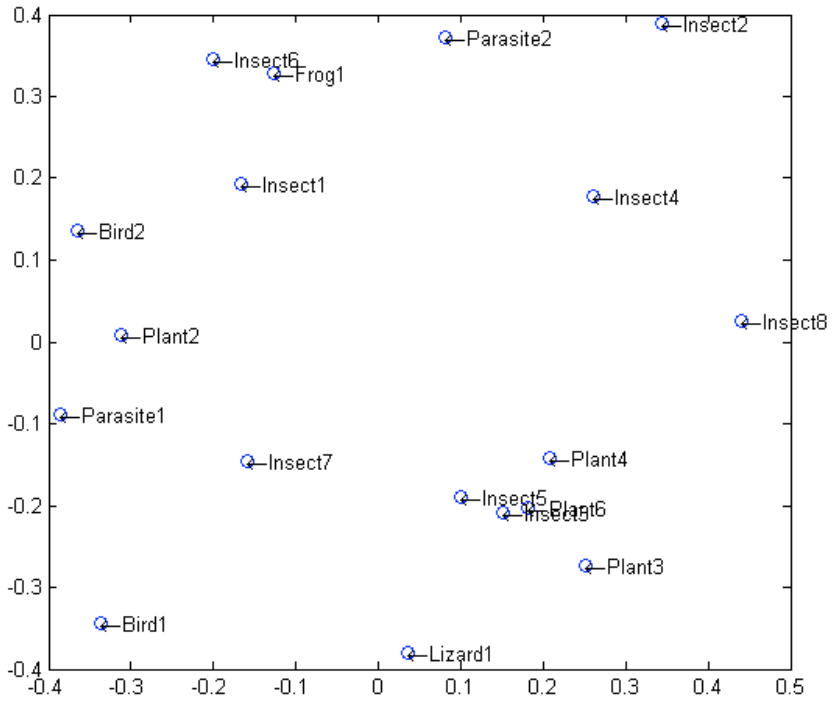

Figure 7: Mutual Information distances projected onto a two-dimensional plot. Axes are in units for the projected space.

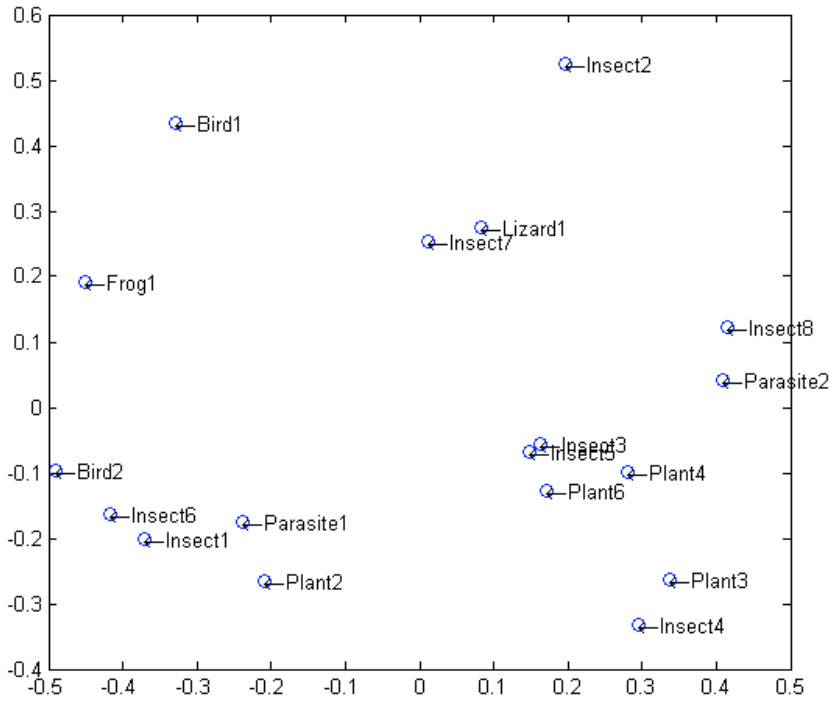

Figure 8: Transfer Entropy distances projected onto a two-dimensional plot. Axes are in units for the projected space.

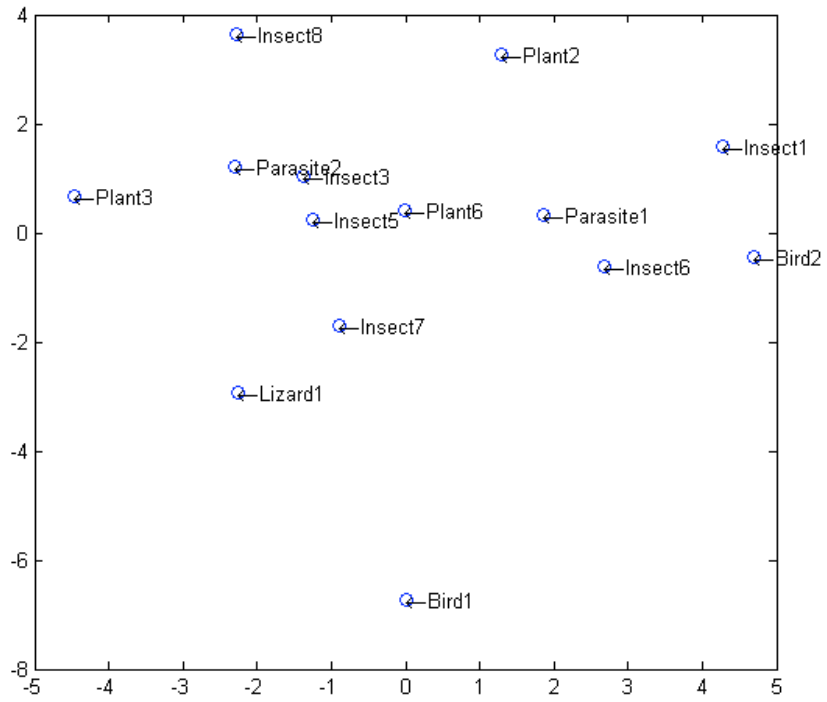

Figure 9: Patch predictor distances projected onto a two-dimensional plot. Axes are in units for the projected space.

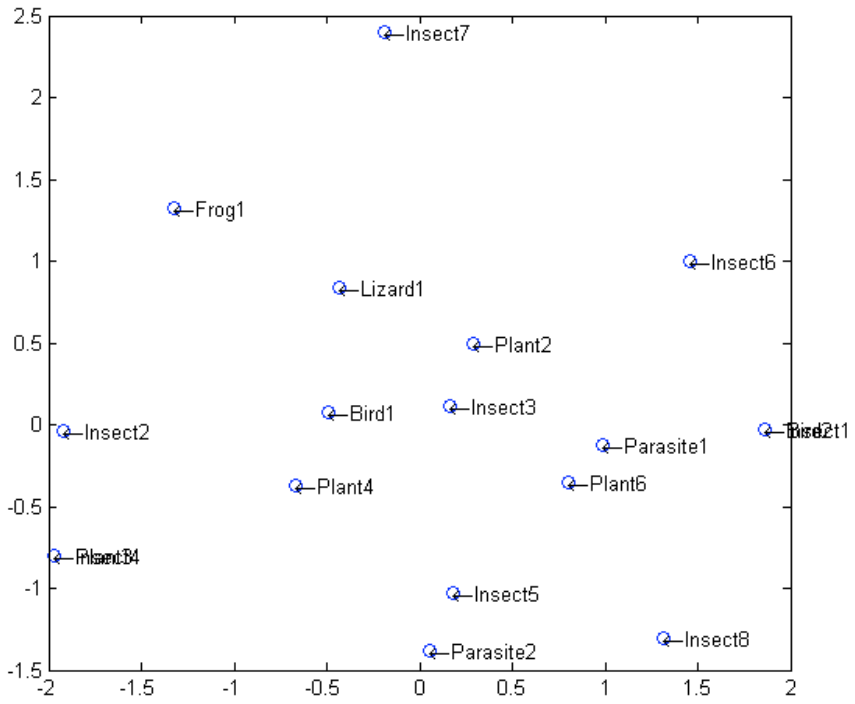

Figure 10: MDS plot shortest path distance of original food web

ever, since there is little evidence to indicate that using the shortest path lengths in the known structure is plausible as a refernce measure.

The latter is supported by a visual inspection of the MDS plots. We can easily identify clusters around "Insect3", "Insect5" and "Plant6" in the mutual information measure, the transfer entropy measure and the patch measure, that do not show in the reference measure.

A more detailed comparison involves analyzing the variation in the number of actual links detected and false links inferred by each technique as a function of a variable threshold distance used to infer the existence of a link. This is analogous to Relative Operating Curve (ROC) plots which are used to evaluate the accuracy of diagnostic systems (e.g. anomaly detection systems for internet servers) in terms of the detection rate and false alarm rate as a function of some sensitivity of the system Swets [1988]. Here, we define analagous parameters for our purposes: the *relationship detection rate* is the proportion of all actual predator–prey relationships which are inferred at the given threshold distance; the false relationship inference rate is the proportion of all non-existent predator–prey relationships which are inferred at the given threshold distance. The ROC plot for each measure is shown in figure 11. This plot demonstrates that the mutual information measure is clearly the least useful distance metric in terms of inferring predator–prey links; in fact it appears no better than a random inference method (due to its close approximation to the diagonal here). At first glance, the transfer entropy and patch predictor distance metrics appear to have similar performance; both demonstrate a more refined ability to distinguish links appropriately. On closer inspection however, we notice that the patch predictor method performs better at low sensitivities (lower detection and false inference rates) while the transfer entropy performs better at higher sensitivities. We believe performance at lower sensitivities is more important, subjectively judging low false inference as more important than high detection because in this region the detection rate is much higher than the false inference rate. As such, we suggest that the patch predictor technique is a marginally better performer on this experiment.

The strength of the patch predictor method appears to lie in detecting the stronger predator–prey links, the proverbial "low hanging fruit" (e.g. at 40 % detection rate, the false inference rate is only 5 %). There is still much room for improvement here though. Perhaps the technique would have performed better given access to larger data sets, or using longer history lengths. Possibly also the technique itself could be improved. While the premise of investigating entanglement in joint models appears sound, the assumption that the joint statistical complexity C_μ would be *smaller* than

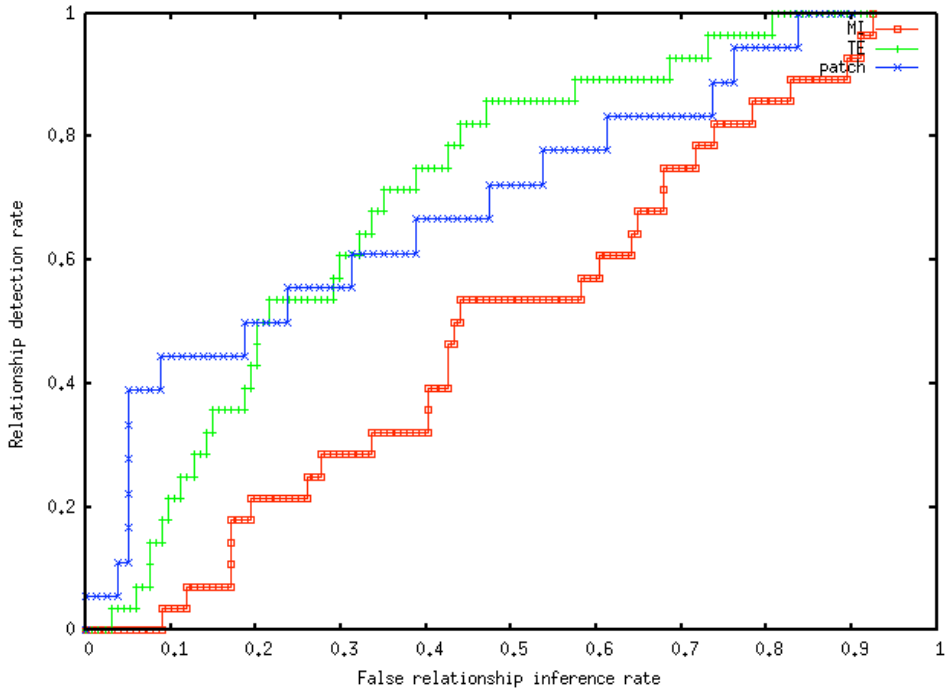

Figure 11: Relative Operating Curve (ROC) plot for predator–prey relationship inference detection rate versus false inference rate for each distance measure (mutual information in red, transfer entropy in green, patch predictor in blue) as a function of variable sensitivity threshold

the sum of the individual statistical complexities for a predator-prey link may not be accurate in all cases. A contrary example predator–prey link is where the predator adds a significant amount of predictive power about the prey, but that information was not relevant to prediction of the predator itself and was not available for analysis of the prey in isolation; here the joint statistical complexity would be larger. We intend to investigate how the measure could be made more comprehensive in future work.

5 Conclusion

Future work shall include reconstructing more complete ϵ-machines for each species using the patch predictors technique to infer the predator–prey links for each species that should be included in their light-cone for the reconstruc-

tion. The patch predictor method shall also be investigated in an attempt to improve its performance.

We think that computational mechanics is one of the most important foundations in complex systems research. It should be more prominent represented in the curriculum of students in particular at the SFI summer school on complex systems in Santa Fe.

Acknowledgments

This work was partially supported by the Santa Fe Institute through NSF Grant No. 0200500 entitled "A Broad Research Program in the Sciences of Complexity." JL acknowledges travel support from the ARC Complex Open Systems Research Network (COSNet) and the CSIRO Complex Systems Science Theme.

References

O. Bochmann. *Probabilistic Approaches in Multi-Agent Systems for Manufacturing Coordination and Control*. PhD thesis, Katholieke Universiteit Leuven, Department of Mechanical Engineering, Leuven, Belgium, June 2005. 1, 2

J. T. Lizier, M. Prokopenko, and A. Y. Zomaya. Local information transfer as a spatiotemporal filter for complex systems. Submitted to Physical Review E, 2007. 1, 2.1

Mathworks. Non-classical multidimensional scaling, 2007. URL `http://www.mathworks.com/products/demos/statistics/mdscaledemo.html`. 4.2

J. M. Nichols. Inferences about information flow and dispersal for spatially extended population systems using time-series data. *Proceedings of the Royal Society B*, 272(1565):871–876, 2005. URL `http://www.journals.royalsoc.ac.uk/content/j5rcv27kvn48hevp/`. 3.2.1

T. Schreiber. Measuring information transfer. *The American Physical Society*, 85(2), 2000. URL `http://prola.aps.org/abstract/PRL/v85/i2/p461_1`. 2.1, 2

C. R. Shalizi. Optimal nonlinear prediction of random fields on networks. *Discrete Mathematics and Theoretical Computer Science*, AB(DMCS):11–30, 2003. URL `arxiv.org/abs/math.PR/0305160`. 2.2

C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. *Statistical Physics*, 104:819–881, 2001. URL `http://www.santafe.edu/~cmg/papers/cmppss.pdf`. 2.2

K. L. Shalizi and C. R. Shalizi. CSSR C++ implementation: An algorithm for building markov models from time series. GNU Public License at `http://cscs.umich.edu/~crshalizi/CSSR/`, 2003. 3.1, 3.2.3

C. E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. The University of Illinois Press, Urbana, 1964. 2.1

J. A. Swets. Measuring the accuracy of diagnostic systems. *Science*, 240 (4857):1285–1293, 1988. 4.2

A. Wagner. How to reconstruct a large genetic network from n gene perturbations in fewer than n^2 easy steps. *Bioinformatics*, 17(12):1183–1197, 2001. URL `http://bioinformatics.oxfordjournals.org/cgi/content/abstract/17/12/1183`. 1

R. Williams, N. Martinez, P. Yoon, I. Yoon, A. Singh, and J. Murakawa. Webs on the web knowledge base, 2007. URL `http://foodwebs.org/wow/webstart/wowKB2.jnlp`. 1, 3, 3