Topological Reversibility and Causality in Feed-Forward Networks

Bernat   Corominas-Murtra
Carlos   Rodriguez-Caso
Joaquin   Goñi
Ricard V. Solé

SANTA FE INSTITUTE

Topological reversibility and causality in feed-forward networks

Bernat Corominas-Murtra[1], Carlos Rodríguez-Caso[1], Joaquin Goñi[2] and Ricard V. Solé[1,3,4]

[1] *ICREA-Complex Systems Lab, Universitat Pompeu Fabra (Parc de Recerca Biomèdica de Barcelona). Dr Aiguader 88, 08003 Barcelona, Spain*

[2] *Functional Neuroimaging Lab. Department of Neurosciences. Center for Applied Medical Research. University of Navarra. Pamplona, Spain*

[3] *Santa Fe Institute, 1399 Hyde Park Road, New Mexico 87501, USA*

[4] *Institut de Biologia Evolutiva. CSIC-UPF. Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Spain.*

Systems whose organization displays causal asymmetry constraints, from evolutionary trees to river basins or transport networks, can be often described in terms of directed paths (causal flows) on a discrete set of arbitrary units -such that states in state spaces, feed-forward neural nets, the evolutionary history of a given collection of events or the chart of computational states visited along a complex computation. Such a set of paths defines a feed-forward, acyclic network. A key problem associated with these systems involves characterizing their intrinsic degree of path reversibility: given an end node in the graph, what is the uncertainty of recovering the process backwards until the origin? Here we propose a novel concept, *topological reversibility*, which is a measure of complexity of the net that rigorously weights such uncertainty in path dependency, quantifying the minimum amount of information required to successfully reverse a causal path. Within the proposed framework we also analytically characterize limit cases for both topologically reversible and maximally entropic structures. The relevance of these measures within the context of evolutionary dynamics is highlighted.

I. INTRODUCTION

Causality is the fundamental principle pervading dynamical processes. Any set of time-correlated events, from the development of an organism to historical changes, defines a feed-forward structure of causal relations captured by a family of complex networks called directed acyclic graphs (DAGs). Their structure has recently attracted the interest of researchers [1–4] since DAGs represent time-ordered processes as well as a broad number of natural and artificial systems. Examples would include simple electronic circuits [5], feed-forward neural [6] and transmission networks [7], river basins [8], or even some food webs and chemical structures [9]. Whereas in some cases the causal interpretation is direct -like, for example, an electronic circuit-, in others -like food webs or river basins- the DAG representation is related to a chart where the dynamics of a given flow is depicted. In these systems, *causality* refers to what is possible if we have, for example a given gradient of potential -like in a river basin- or energy -like in a food web of a given ecosystem.

A paradigmatic example of a causal structure is the chart of the relations among states followed by a computational process through time. Intimately linked to the topology of such computational chart, a fundamental feature of computations is its degree of *logical reversibility* [10, 11]. Indeed, it is said that a process is *logically reversible* when, if reversing the flow of causality, i.e. going backwards from the computational outputs to their inputs, we can unambiguously recover the causal structure of the process. Roughly speaking, if we have a computer performing a function $g : \mathbb{N} \to \mathbb{N}$ and we can unambiguously determine the input u from the only knowledge of the value $v = g(u)$, (i.e, g is a bijection) we say that the computation is logically reversible. Otherwise, if there is uncertainty in determining u from the only knowledge of v, we say that the computation is *logically irreversible*, and thus, additional information is needed to successfully reconstruct a given computational path. Analogously, the potential scenarios emerging from an evolutionary process raise similar questions. Within evolutionary biology, a relevant problem is how predictable is evolutionary dynamics. In particular, it has been asked what would be the result of going backwards and "re-playing the tape of evolution" [12, 13]. Since this question pervades the problem of how uncertain or predictable is a given evolutionary path, it seems desirable to actually provide a foundational framework. In particular, the picture of evolutionary dynamics based on climbing fitness landscapes [14] allows us to depict adaptive changes as directed paths that can be in principle mapped into a DAG. This would be the case for discrete fitness landscapes associated to sequence spaces, which can be climbed in an irreversible fashion through adaptive walks [15, 16].

In this paper, we analytically extend the concept of logical reversibility to the study of any causal structure having no cyclic topologies, thereby defining a broader concept to be named *topological reversibility*. Whereas thermodynamical irreversibility implies thermodynamical entropy production [17, 18], topological irreversibility implies statistical entropy *production*. In general, we will say that a DAG is *topologically reversible* if we can unambiguously recover a path going backwards from any element to the origin. Genealogies and phylogenies are examples of tree-like structures where a chronological order can be established among the events and an unam-
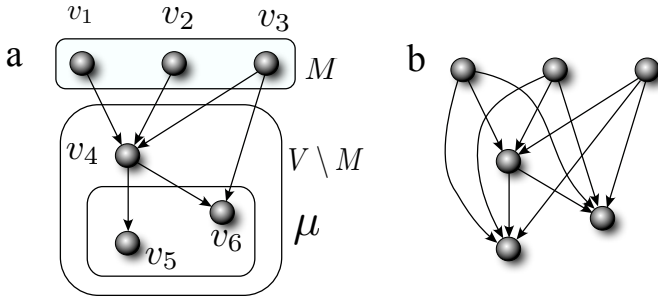
FIG. 1: Some illustrative DAGs. A topologically irreversible DAG $\mathcal{G}(V, E)$, where M denotes the set of maximals, μ the set of minimals and the $V \setminus M$ set the set of non-maximals (a). The respective transitive closure, $\mathcal{T}(\mathcal{G})$ is shown in (b) -see text.

biguous reconstruction of the lineage can be performed for every element of the graph [19]. Following this argument, we will label a graph as *topologically irreversible* when some uncertainty is observed in the reconstruction of trajectories.

As shown below, the entropy presented here weigths the extra amount of information that would be required to recover the causal flow backwards. In this way, it can be understood as a measure of complexity of the paths defined by the net. The so-called *path diversity* or [20] or other complexity measures based on path heterogeneity [21], [22] can be somehow understood as conceptual precursors of the measure we define here. Other information measures have been defined in the study of complex networks [23–29], although such measures accounted for connectivity correlations [24, 25, 27, 28] or were used to characterize a Gibbsian formulation of the statistical mechanics of complex networks [23]. We finally note that the starting point of our formalism resembles the classical theory of Bayesian networks. However, the particular treatment of reversibility proposed here is qualitatively different from the concept of uncertainty used in such a framework and closer to the one described in [26].

The paper is organized as follows: In section II we provide the basic concepts underlying our analytical derivations. Section III provides the general mathematical definition of topological reversibility and the general expression for the average uncertainty associated to the reversion of the causal flow. This is consistently derived from the properties of the adjacency matrix. In section IV we consider two limit cases, finding the exact analytic form for their entropies and predicting the most uncertain configuration. Finally, in section V we outline the generality and relevance of our results in terms of characterizing DAG structure.

II. THEORETICAL BACKGROUND

The theoretical roots of this paper stem from fundamental notions of directed graph theory [30, 31], ordered set theory [32, 33] and information theory [34–37]. Specifically, we make use of Shannon's entropy which, as originally defined, quantifies the uncertainty associated to certain collections of random events [34, 36]. In our framework, the entropy in a given feed-forward graph measures the uncertainty in reversing the causal flow depicted by the arrows[45].

A. Directed graphs and orderings

Let $\mathcal{G}(V, E)$ be a directed graph, being $V = \{v_1, ..., v_n\}$, $|V| = n$, the set of nodes, and $E = \{\langle v_k, v_i \rangle, ..., \langle v_j, v_l \rangle\}$ the set of edges. The ordered pair notation, $\langle v_k, v_i \rangle$, implies that there is an arrow in the following direction:

$$v_k \rightarrow v_i.$$

Given a node $v_i \in V$, the number of outgoing links, to be written as $k_{out}(v_i)$, is called the *out-degree* of v_i and the number of ingoing links of v_i is called the *in-degree* of v_i, written as $k_{in}(v_i)$. The *adjacency matrix* of a given graph $\mathcal{G}$, $\mathbf{A}(\mathcal{G})$ is defined as $A_{ij}(\mathcal{G}) = 1 \leftrightarrow \langle v_i, v_j \rangle \in E$; and $A_{ij}(\mathcal{G}) = 0$ otherwise. Through the adjacency matrix, k_{in} and k_{out} are computed as

$$k_{in}(v_i) = \sum_{j \leq n} A_{ji}(\mathcal{G}); \quad k_{out}(v_i) = \sum_{j \leq n} A_{ij}(\mathcal{G}).$$

Furthermore, we will use the known relation between the k-th power of the adjacency matrix and the number of paths of length k going from a given node v_i to a given node v_j Specifically,

$$(\mathbf{A}(\mathcal{G}))_{ij}^k = (\overbrace{\mathbf{A}(\mathcal{G}) \times ... \times \mathbf{A}(\mathcal{G})}^{k \text{ times}})_{ij}$$

is the number of paths of length k going from node v_i to node v_j [31].

A *feed-forward* or *directed acyclic graph* is a directed graph characterized by the absence of cycles: If there is a *directed path* from v_i to v_k (i.e., there is a finite sequence $\langle v_i, v_j \rangle, \langle v_j, v_l \rangle, \langle v_l, v_s \rangle, ..., \langle v_m, v_k \rangle \in E$) then, there is no directed path from v_k to v_i. Conversely, the matrix $\mathbf{A}^T(\mathcal{G})$ depicts a DAG with the same underlying structure but having all the arrows (and thus, the causal flow) inverted. Given its acyclic nature, one can find a finite value $L(\mathcal{G})$ as follows:

$$L(\mathcal{G}) = \max\{k : (\exists v_i, v_j \in V : (\mathbf{A}(\mathcal{G}))_{ij}^k \neq 0)\}. \quad (1)$$

It is easy to see that $L(\mathcal{G})$ is the length of the longest path of the graph. The existence of such $L(\mathcal{G})$ can be seen as a test for acyclicity. However, the use of leaf-removal algorithms [38, 39], i.e. the iterative pruning

of nodes without outgoing links, is by far more suitable than the above method, in terms of computational costs. In a DAG, a leaf-removal algorithm removes completely the graph in a finite number of iterations, specifically, in $L(\mathcal{G})$ iterations -see eq. (1).

Now we study the interplay between DAGs and order relations. Borrowing concepts from order theory [33], we define the following set:

$$M = \{v_i \in V : k_{in}(v_i) = 0\},$$

to be named the set of *maximal nodes* of $\mathcal{G}$, by which $|M| = m$. The set of all paths $\pi_1, ..., \pi_s$, $s \geq |E|$, from M to any node $v_i \in V \setminus M$ is indicated as $\Pi(\mathcal{G})$. Given a node $v_i \in V \setminus M$, the set of all paths from M to v_i is written as $\Pi(v_i) \subseteq \Pi(\mathcal{G})$. Furthermore, we will define the set $v(\pi_k)$ as the set of all nodes participating in this path, except the maximal one. Additionally, one can define the set of nodes with $k_{out} = 0$ as the set of *minimal nodes* of $\mathcal{G}$, to be named μ. Notice that the absence of cycles implies that $m \geq 1$ and that the set of minimals μ must also contain at least one element -see fig. (1a).

Attending to the node relations depicted by the arrows, and due to the acyclicity property, at least one node ordering can be defined, establishing a natural link between order theory and DAGs. This order is achieved by labeling all the nodes with sequential natural numbers and obtaining a configuration such that:

$$(\forall \langle v_i, v_j \rangle \in E)(i < j). \tag{2}$$

Accordingly, DAGs are *ordered graphs* [2]. However, as order relations imply *transitivity*, it is not the DAG but its transitive closure what properly defines the order relation among the elements of V. The *transitive closure* of $\mathcal{G}$ (see fig. 1b), to be written as $T(\mathcal{G}) = (V_T, E_T)$ is defined as follows: Any pair of nodes $v_i, v_k \in V$ by which there is at least one (directed) path going from v_i to v_k are connected through a link $\langle v_i, v_k \rangle$ in $T(\mathcal{G})$. In this framework, for a given number of maximal nodes, in the transitive closure the addition of a link either creates a cycle or destroys a maximal or minimal node. If the pairs defining the set of links of $T(\mathcal{G})$ are conceived as the elements of a set relation $E_T \subset V \times V$, such a relation satisfies the following three properties:

i) $\nexists \langle v_k, v_k \rangle$,
ii) $(\langle v_i, v_k \rangle \in E_T) \Rightarrow (\langle v_k, v_i \rangle \notin E_T)$,
iii) $(\langle v_i, v_k \rangle \in E_T \wedge \langle v_k, v_j \rangle \in E_T) \Rightarrow (\langle v_i, v_j \rangle \in E_T)$.

The DAG definition implies that E directly satisfies the two first conditions whilst the third one (transitivity) is only warranted for E_T. Thus, only E_T holds all requirements to be an *order relation*, specifically, a *strict partial order*. The transitive closure of a given DAG can be obtained by means of the so-called *Warshall's* algorithm [31].

Finally, a subgraph $\mathcal{F}(V_{\mathcal{F}}, E_{\mathcal{F}}) \subseteq \mathcal{G}$ is said to be *linearly ordered* or *totally ordered* provided that for all pairs of nodes $v_i, v_k \in V_{\mathcal{F}}$ such that $k < i$, then

$$\langle v_k, v_i \rangle \in E_{\mathcal{F}}. \tag{3}$$

Let us notice that if we understand $E_{\mathcal{F}}$ as a set relation $E_{\mathcal{F}} \subset V_{\mathcal{F}} \times V_{\mathcal{F}}$, $E_{\mathcal{F}}$ is a *strict linear order*. If $\mathcal{G}$ is linearly ordered and $\mathcal{W} \subset \mathcal{G}$, we refer to $\mathcal{G}$ as a *topological sort* of $\mathcal{W}$ [31].

B. Uncertainty

According to classical information theory [34–37], let us consider a system S with n possible states, whose occurrences are governed by a random variable X with an associated probability mass function formed by $p_1, ..., p_n$. According to the standard formalization, the *uncertainty* or *entropy* associated to X, to be written as $H(X)$, is:

$$H(X) = -\sum_{i \leq n} p_i \log p_i,$$

which is actually an average of $\log(1/p(X))$ among all events of S, namely, $H(X) = \langle \log(1/p(X)) \rangle$, where $\langle ... \rangle$ is the *expectation* or average of the random quantity between parentheses. As a concave function, the entropy satisfies the so-called *Jensen's inequality* [35], which reads:

$$\left\langle \log \frac{1}{p(X)} \right\rangle \leq \log \left\langle \frac{1}{p(X)} \right\rangle \leq \log n, \tag{4}$$

The maximum value $\log n$ is achieved for $p_i = 1/n$ for all $i = (1, ..., n)$. Jensen's inequality provides an upper bound on the entropy that will be used below. Analogously, we can define the *conditional entropy*. Given another system S' containing n' values or choices, whose behavior is governed by a random variable Y, let $\mathbb{P}(s'_i|s_j)$ be the conditional probability of obtaining $Y = s'_i \in S'$ if we already know $X = s_j \in S$. Then, the conditional entropy of Y from X, to be written as $H(Y|X)$, is defined as:

$$H(Y|X) = -\sum_{j \leq n} p_j \sum_{i \leq n'} \mathbb{P}(s'_i|s_j) \log \mathbb{P}(s'_i|s_j). \tag{5}$$

which is typically interpreted as a noise term in information theory. Such a noise term can be understood as the minimum amount of extra bits needed to unambiguously determine the input set from the only knowledge of the output set. This will be the key quantity of our paper, for it accounts for the *dissipation* of information in a given process.

III. TOPOLOGICAL REVERSIBILITY AND ENTROPY

Let us imagine that a node $v_i \in V \setminus M$ of a given DAG $\mathcal{G}$, receives the visit of a random walker that follows the
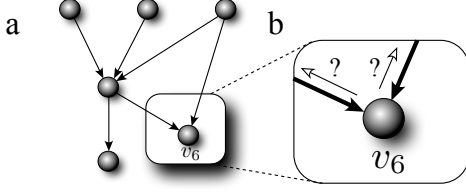
FIG. 2: Uncertainty in the reversal of causal flows in a DAG. Notice that more than one pathway, with more or less probability to be chosen, connect maximals from each terminal (a). Given a node (v_6) receiving two inputs, we consider two different alternatives to go backwards. The uncertainty in this particular case is obtained by computing $h_{loc}(v_i)$ from eq. (III B), i.e., $h_{loc}(v_6) = \log 2$ assuming equiprobability in the selection (b)

.

flow chart depicted by the DAG. We only know that it began its walk at a given maximal node and it followed a downstream random path attending to the directions of the arrows to reach the node v_i. What is the uncertainty associated to the followed path? In other words, what is the amount of information we need, on average, to successfully perform the backward process?

A. The definition of entropy

As we mentioned above, the starting point of our derivation is somewhat close to the treatment of Bayesian networks [40]. In our approach, the first task is to define the probability to follow a given path $\pi_k \in \Pi(v_i)$ when reversing the process. Let $v(\pi_k)$ be the set of nodes participating in the path π_k except the maximal ones. Maximal nodes are not included in this set because they are the ends of the path of the reversal process. The probability to chose such a path from node v_i by making a random decision at every crossing when reversing the causal flow will be:

$$\mathbb{P}(\pi_k|v_i) = \prod_{v_j \in v(\pi_k)} \frac{1}{k_{in}(v_j)}. \tag{6}$$

Consistently:

$$\sum_{\pi_k \in \Pi(v_i)} \left(\prod_{v_j \in v(\pi_k)} \frac{1}{k_{in}(v_j)} \right) = 1.$$

As $\mathbb{P}$ is a probability distribution, we can compute the uncertainty associated to a reversal of the causal flow, starting the reversion process from a given node $v_i \in$

$V \setminus M$, to be written as $h(v_i)$:

$$h(v_i) = - \sum_{\pi_k \in \Pi(v_i)} \mathbb{P}(\pi_k|v_i) \log \mathbb{P}(\pi_k|v_i)$$

The overall uncertainty of $\mathcal{G}$, written as $H(\mathcal{G})$, is computed by averaging h over all non-maximal nodes, i.e:

$$H(\mathcal{G}) = - \sum_{v_i \in V \setminus M} p(v_i) \sum_{\pi_k \in \Pi(v_i)} \mathbb{P}(\pi_k|v_i) \log \mathbb{P}(\pi_k|v_i)$$
$$= \sum_{v_i \in V \setminus M} p(v_i) h(v_i). \tag{7}$$

B. The transition matrix Φ and its relation to the adjacency matrix

The main combinatorial object of our approach is not the adjacency matrix but instead a mathematical representation of the probability to visit a node $v_i \in V \setminus M$ starting the backward flow from a given, different node $v_k \in V \setminus M$ regardless the distance separating them. As we shall see, this combinatorial information can be encoded in a matrix, to be named *transition matrix* Φ and we can explicitly obtain it from $\mathbf{A}(\mathcal{G})$. We begin by defining

$$V(\Pi(v_j)) \equiv \bigcup_{\pi_k \in \Pi(v_j)} v(\pi_k),$$

and we can see that:

$$h(v_i) = - \sum_{\pi_k \in \Pi(v_i)} \mathbb{P}(\pi_k|v_i) \log \mathbb{P}(\pi_k|v_i)$$
$$= \sum_{\pi_k \in \Pi(v_i)} \left[\sum_{v_j \in v(\pi_k)} \mathbb{P}(\pi_k|v_i) \log(k_{in}(v_j)) \right].$$

At this point, we notice that we obtained a factorization of the particular contribution of the entropy of every node, $\log(k_{in}(v_j))$ and the probability for this node to participate in a given causal path starting at M and ending at v_i, namely $\sum_{v_j \in v(\pi_k)} \mathbb{P}(\pi_k|v_i)$. Let us thus define ϕ_{ij} as the coefficients of a $(n-m) \times (n-m)$ matrix $\Phi(\mathcal{G}) = [\phi_{ij}(\mathcal{G})]$, i.e. our *transition matrix* $\Phi(\mathcal{G})$:

$$\phi_{ij}(\mathcal{G}) = \sum_{\pi_k : v_j \in v(\pi_k)} \mathbb{P}(\pi_k|v_i).$$

For the moment, we will not develop its explicit form. Therefore:

$$= \sum_{v_j \in V(\Pi_i)} \log(k_{in}(v_j)) \left[\sum_{\pi_k : v_j \in v(\pi_k)} \mathbb{P}(\pi_k|v_i) \right]$$
$$= \sum_{v_k \in V \setminus M} \phi_{ik}(\mathcal{G}) h_{loc}(v_k). \tag{8}$$

Let us explain eq. (8) and its consequences. First we define $h_{loc}(v_i)$ as:

$$h_{loc}(v_i) = \log(k_{in}(v_i)).,$$

where L indicates the amount of *local* entropy introduced in a given node when performing the reversion process - see fig (2). Thereby, it is the amount of information needed to properly reverse the flow backwards when a bifurcation point is reached having k_{in} possible choices.

Now we derive the general expression for Φ. The derivation allows us to obtain a consistent mathematical definition of the transition matrix in terms of $\mathbf{A}(\mathcal{G})$. We first notice two important facts linking paths and the powers of the adjacency matrix that are only generically valid in DAG-like networks. First, we observe that:

$$|\Pi(v_i)| = \sum_{j \leq L(\mathcal{G})} \sum_{l:v_l \in M} (\mathbf{A}^T(\mathcal{G}))_{il}^j, \qquad (9)$$

being $L(\mathcal{G})$ the length of the longest path of the graph as defined by (1). Analogously, the number of paths of $\Pi(v_i)$ crossing v_k, to be written as α_{ik} is:

$$\alpha_{ik} \equiv |\{\pi_j \in \Pi(v_i) : v_k \in v_i(\pi_j)\}|$$
$$= \sum_{\ell \leq L(\mathcal{G})} \left(\mathbf{A}^T(\mathcal{G})\right)_{ik}^{\ell}.$$

The above quantities provide the number of paths. To compute the probability to reach a given node, we have to take into account the probability to follow a given path containing such a node, defined in (6). To rigorously connect it to the adjacency matrix, we first define an auxiliary, $(n-m) \times (n-m)$ matrix $\mathbf{B}(\mathcal{G})$, namely:

$$B(\mathcal{G})_{ij} = (A_{ij}(\mathcal{G})) \left(\sum_{j \leq n} A_{ij}(\mathcal{G})\right)^{-1} = \frac{A_{ij}(\mathcal{G})}{k_{in}(v_i)},$$

where $v_i, v_j \in V \setminus M$ and, as defined in previous sections, $n = |V|$, $m = |M|$. From this definition, we obtain the explicit dependency of Φ from the adjacency matrix, namely[46],

$$\phi_{ij}(\mathcal{G}) = \sum_{k \leq L(\mathcal{G})} \left(\mathbf{B}^T(\mathcal{G})\right)_{ij}^k. \qquad (10)$$

and accordingly, we have

$$\phi_{ii}(\mathcal{G}) = \left(\mathbf{B}^T(\mathcal{G})\right)_{ii}^0 = 1. \qquad (11)$$

It is worth to mention that $\Phi(\mathcal{G})$ resembles the transition matrix related to the concept of *information mobility* [26]. In the general case of non-directed graphs, one can assume the presence of paths of arbitrary length, which leads (using a correction factor tied to the length of the path) up to an asymptotic form of the transition matrix in terms of the exponential of the adjacency matrix. However, the intrinsic finite nature of the paths in a given DAG makes the above asymptotic treatment non viable.

C. The general form of the Entropy

Let us now define the overall entropy in a compact form, only depending on the adjacency matrix of the graph. From eqs. (5, 7, 8), we obtain

$$H(\mathcal{G}) = \sum_{v_i \in V \setminus M} p(v_i) \sum_{v_k \in V \setminus M} \phi_{ik}(\mathcal{G}) h_{loc}(v_k). \qquad (12)$$

This is the central equation of this paper. This measure quantifies the additional information (other than topological one) to properly reverse the causal flow. We observe that this expression is a *noise* term within standard information theory [34]. In this equation we have been able to decouple the combinatorial term associated to the multiplicity of paths at one hand, and the particular contribution to the overall uncertainty of every node, at the other hand. The former is fulfilled by the matrix Φ, which encodes combinatorial properties of the system, and how they influence in the computation of the entropies. The latter is obtained from the set of local entropies $h_{loc}(v_1), ..., h_{loc}(v_{n-m})$. These terms account for the contribution of local topology -i.e. the uncertainty when choosing an incoming link at the node level in the reversion of the causal flow- to the overall entropy. This uncoupling is a consequence of the extensive property of the entropy and, putting aside its conceptual interest, simplifies all derivations related to the uncertainties. This general expression of the entropy can be simplified if we assume that $\forall v_i \in V \setminus M$, $p(v_i) = 1/(n-m)$. Therefore, by defining

$$Q(\mathcal{G}) = \sum_{v_i \in V \setminus M} \sum_{v_k \in V \setminus M} \phi_{ik}(\mathcal{G}) h_{loc}(v_k) \qquad (13)$$

and thus $H(\mathcal{G})$ is expressed as:

$$H(\mathcal{G}) = \frac{1}{n-m} Q(\mathcal{G}) \qquad (14)$$

Finally, we recall that the above entropy is bounded by Jensen's inequality (4) to equation (8)i.e.,

$$H(\mathcal{G}) \leq \frac{1}{n-m} \sum_{v_i \in V \setminus M} \log(|\Pi(v_i)|). \qquad (15)$$

Notice that the quantity on the right side of eq. (15) is the uncertainty obtained by considering all paths from M to v_i equally likely to occur.

D. Topological reversibility

Having defined an appropriate and well grounded entropy measure, now we can discuss the meaning of *topological (ir)reversibility*. Let us first make a qualitative link with standard theory of irreversible thermodynamics, where irreversibility is tied to the parameter of entropy production σ^s in the entropy balance equation [18].
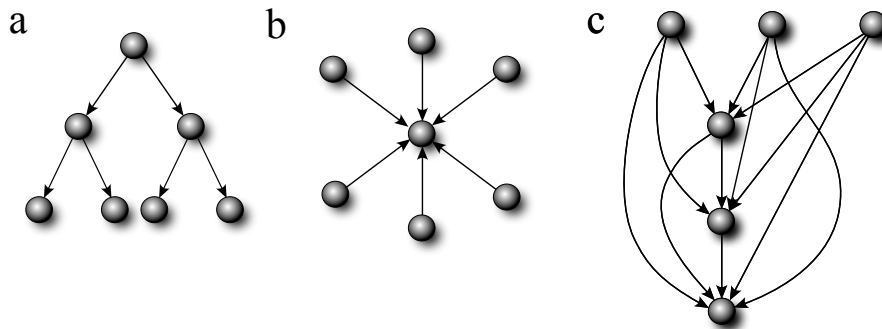
FIG. 3: Limit cases of graphs. (a) A topological reversible structure featured by a tree DAG structure, $H(\mathcal{G}) = 0$. (b) With the same number of nodes and links, its opposite structure is a topologically irreversible DAG featured by a star DAG with $m = n - 1$. Notice that for a star graph $H(\mathcal{G}) = \log(n-1)$ where $n = 7$ in this particular case. (c) Graphical representation of the most entropic graph having 3 maximal nodes and 3 minimal nodes. Notice that the set $V \setminus M$ is linearly ordered and that every maximal is connected to every non-maximal -see text and Appendix A.

Here, $\sigma^s = 0$ depicts *thermodynamically reversible* processes, whereas $\sigma^s > 0$ appears in *irreversible* processes [17, 18]. Irreversibility is rooted in the impossibility of reversing the process without generating a negative amount of entropy, which contradicts to the second law of thermodynamics. Consistently, we will call *topologically reversible* those DAG structures such that

$$H(\mathcal{G}) = 0.$$

In those structures (they belong to the set of trees, as we shall see in the following section) no ambiguity arises when performing the reversion process. On the contrary, a given DAG by which

$$H(\mathcal{G}) > 0$$

will be referred to as *topologically irreversible*. DAGs having $H(\mathcal{G}) > 0$ display some degree of uncertainty taking the causal flow backwards, since the reversion process is subject to some random inevitable decisions. In these cases, $H(\mathcal{G})$ is the average of the amount of extra information needed to successfully perform the process backwards. Similarly, the successful reversion of a thermodynamically irreversible process would imply the (irreversible) addition of external energy, or that the reversion of a logically irreversible computation requires an extra amount of external information to solve the ambiguity arising in rewinding the chain of computations. In this context, for example, reversible computation is defined by considering a system of storage of history of the computational process [10]. Furthermore, we observe that, roughly speaking, we can associate the logical (ir)reversibility of a computational process to the topological (ir)reversibility of its DAG representation. In our study, the adjective *topological* arises from the fact that we only use topological information to compute the uncertainty. Thus, we deliberately neglect the active role that a given node can play as, for example, a processing

unit, or the different weights of the paths. However, it is worth to mention that entropy can be generalized for DAGs where links are weighted by a probability to be chosen in the process of reaching the maximal.

IV. LIMIT CASES: MAXIMUM AND MINIMUM UNCERTAINTY

Let us illustrate our previous results by exploring two limit cases, namely DAGs having zero or maximal uncertainty. In this section we identify those feed-forward structures which, containing n nodes and without a predefined number of links, minimize or maximize the above uncertainties. The minimum uncertainties are obtained when the graph $\mathcal{G}$ is a special kind of tree -a topologically reversible structure-, to be described below. Afterwards, we summarize the main properties of the graph configuration displaying maximum entropy. The detailed derivations of the latter case are slightly sophisticated and are provided in the Appendix A.

A. Zero Uncertainty: Trees

Imagine a random walker exploring a (directed) tree containing only a single maximal (fig. 3a). From such a maximal node, there exists only one path to a given node. In the evolutionary context, a single ancestor is at the root of all evolutionary tree [41]. Thus, the process of recovering the history of the random walker up to its initial condition is completely deterministic, and no uncertainty can be associated to it -in purely topological terms. Formally, we recognize two defining features on trees, namely:

- $m = 1$

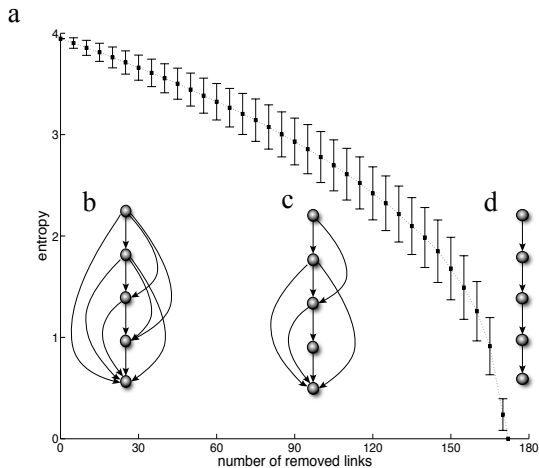- $(\forall v_i \in V \setminus M)(k_{in}(v_i) = 1)$.

FIG. 4: a) Evolution of the Entropy (in bits) starting from a linearly ordered DAG having $n = 25$ nodes and one maximal, i.e., $\tilde{\mathcal{G}}_{25}(1)$, and then successively removing a randomly chosen link until we obtain a linear chain, a topologically reversible structure, thereby having zero entropy. Every point of the graph is obtained by evaluating the entropies of the members the ensemble $g(25, 1)$ having a defined number of links -the search is performed by computing 500 replicas. We appreciate the decreasing trend as predicted -see text. In b) we show a qualitative picture of the kind of graph we find as starting point of the numerical exploration. c) Shows how it looks like an intermediate member and d) the final one.

.

We thus conclude that there is no uncertainty in recovering the flow, since the two reported properties are enough to conclude that there is 1 and only 1 path to go from M to any $v_i \in V \setminus M$. This agrees with the intuitive idea that trees are perfect hierarchical structures. This result complements the more standard scenario of the forward, downstream scenario paths followed by a random walker on a tree [19]. It is worth noting that evolutionary trees, particularly in unicellular organisms, have been found to be a poor representation of the actual evolutionary process [42, 43].

B. Maximum Uncertainty

Now we shall concern ourselves with the maximally entropic scenario(s). We refer the reader interested on the mathematical details to the derivations contained in the Appendix A. We observe that, intuitively, a star star DAG having n nodes, being $m = n - 1$ the maximal ones can be understood as opposite to a tree, in terms of $H(\mathcal{G})$. Specifically, it is easy to see that, whereas a tree $\mathcal{G}_T$ having n nodes displays $H(\mathcal{G}_T) = 0$, a star DAG, $\mathcal{G}_\star$, $m = n - 1$ maximal nodes has an entropy like:

$$H(\mathcal{G}_\star) = \log(n - 1).$$

However, the combinatorial nature of the defined entropy enables to more complex, non intuitive structures to display higher values. Indeed, let $g(n, m)$ be the ensemble of all connected DAG structures having n nodes, being m of them the maximal ones. It can be shown -see Appendix A- that the most entropic graph of the ensemble $g(n, m)$ -to be referred as $\tilde{\mathcal{G}}_n(m)$- is the graph obtained applying the following rules:

1. For all pair $\{v_i, v_k\}$ such that $v_i \in M$ and $v_k \in V \setminus M$ generate the link

$$\langle v_i, v_k \rangle$$

2. For every pair $\{v_j, v_l\}$ such that $v_j, v_l \in V \setminus M$ and such that $j < l$ generate the link

$$\langle v_j, v_l \rangle.$$

The exact expression for the entropy of the above constructed graph is -see Appendix A:

$$H(\tilde{\mathcal{G}}_n(m)) = \frac{n}{n - m} \sum_{i \leq n - m} \frac{\log(m + i - 1)}{m + i}$$

Finally, we emphasize that it can be shown that, for $n \gg 1$, the most entropic DAG having n nodes and having no constraints related to the number of maximal nodes or total connections is $\tilde{\mathcal{G}}_n(2)$, i.e., $\forall k \neq 2$, $k < n$:

$$H(\tilde{\mathcal{G}}_n(2)) > H(\tilde{\mathcal{G}}_n(k)),$$

and, therefore, since $\forall \mathcal{G} \in g(n, m)$, $H(\tilde{\mathcal{G}}_n(m)) > H(\mathcal{G})$, $\tilde{\mathcal{G}}_n(2)$ displays the absolute maximum of the entropy when having n possible nodes and a DAG-like structure. The active role played by the combinatorial pattern and the number of links leads to this somehow unexpected result. It should be noted that intermediate configurations can also display high entropies -see figure (4). The general problem, i.e., to find the most entropic configuration having a fixed number of links and nodes is not addressed here. Instead of it, we made a numerical exploration of the evolution of the entropies starting from a given $\tilde{\mathcal{G}}_n(m)$ and removing links at random until we have a chain -a perfect reversible structure- attached to a given number of maximals -see figure (4). We observe that, according to our derivations -see lemma 1 in the Appendix A- the removal of links result in a net reduction of uncertainty.

V. DISCUSSION

In this paper we address the problem of quantifying path dependencies using the DAG metaphor. To this goal, we introduce the concept of topological reversibility as a fundamental feature of causal processes that can be depicted by a DAG structure. The intuitive definition is rather simple: A system formed by an aggregation of causal processes is topologically reversible if we

can recover all causal paths with no other information than the one provided by the topology. If graph topology induces some kind of ambiguity in the backward process, the graph is said to be topologically irreversible, and additional information is needed to build the backward flows.

We provided the analytical form of the uncertainty (the amount of extra information needed) arising in the reversion process by uncoupling the combinatorial information encoded by the graph structure from the contributions of the local connectivity patterns of individual nodes, as depicted in eqs. (13,14). It is worth noting that all our results are derived from just two basic concepts: The adjacency matrix of the graph and the definition of entropy. Furthermore, we offer a constructive derivation of the two limit cases, namely trees (as the reversible ones), and linear ordered graphs (having two maximal nodes) as the most uncertain ones.

According to our results, only a tree DAG is topologically reversible. However, beyond this particular case, the quantification of topological irreversibility by using the entropy proposed here could provide insights in the characterization of more general feed forward systems. An illustrative case-study can be found precisely in biological evolution. The standard view of the tree of life involves a directional, upward time-arrow where the genetic structure of a given species (its genome) derives from some ancestor after splitting (speciation) events. One would think that this classical but too simplistic view of evolution as a tree gives a topologically reversible lineage of genes, changing by mutations and passing from the initial ancestor to current species in a vertical inheritance. However, it has been recently evidenced that the so-called horizontal gene transfer among unrelated species may have had a deep impact in the evolution and diversification in microbes [43]. According to this genetic mechanism the tree-like picture and thus the logical/topological reversibility is broken by the presence of cross-links between *brother* species. Under the light of these evidences, tree-based phylogenies become unrealistic. In this context, our theoretical approach provides a suitable framework for the characterization of the logical irreversibility of biological evolution and, in general, for any process where time or energy dissipate impose a feed-forward chart of events. Further research in this topic will contribute to understand the causal structure of evolutionary processes.

Appendix A: Derivation of the most Entropic structure

In this mathematical appendix we construct in detail the most entropic DAG.

1. Construction of $\tilde{\mathcal{G}}$: The linear ordering in $V \setminus M$.

Let $\mathcal{G}$ be a feed-forward organized graph containing n nodes, where m of them are maximal. Since for the entropy computation all nodes become indistinguishable, let $g(m,n)$ be the ensemble of all different possible feed-forward configurations containing n nodes, where m of them are maximal. We are looking for a graph, to be written as

$$\tilde{\mathcal{G}}_n(m) \in g(n,m)$$

such that $\forall \mathcal{G}_i \in g(m,n)$:

$$\mathcal{G}_i \subseteq \tilde{\mathcal{G}}_n(m),$$

i.e., a graph containing all possible links, preserving the number of maximal nodes and the property of being a DAG. This implies, as defined in section II A, eq. (3), that we must add links to the set $V \setminus M$ until it becomes *linearly ordered*, attending to a labeling of nodes which respect the ordering depicted by the feed-forward graph -see figure (3c). Once we have the set of nodes $V \setminus M$ linearly ordered, we proceed to generate a link from any node $v_i \in M$ to every node $v_k \in V \setminus M$. We thus obtain a feed forward graph containing m maximal nodes and only 1 minimal node. In the above constructed graph, any new link creates a cycle or destroys a maximal vertex. For the sake of clarity we will differentiate the labeling of M and $V \setminus M$ when working with $\tilde{\mathcal{G}}_m(n)$. Specifically, nodes $v_i \in V \setminus M$ will be labeled sequentially from 1 to $n - m$ respecting the ordering defined in eq. (2). This labeling will be widely used in the forthcoming sections. Furthermore, we recall that no special labeling other than different natural numbers is needed for $v_k \in M$, since there will be no ambiguous situations.

As an example, we explicitly write the adjacency matrix of $\tilde{\mathcal{G}}_6(3)$, depicted in figure...

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Now we are ready to demonstrate the first two lemmas of the appendix.

Lemma 1: *Given two fixed values of m and n and two DAGs, $\tilde{\mathcal{G}}_n(m)$ and $\mathcal{G} \in g(n,m)$, $\mathcal{G} \neq \tilde{\mathcal{G}}_n(m)$, the following inequality holds[47]:*

$$H(\tilde{\mathcal{G}}_n(m)) > H(\mathcal{G}).$$

Proof. Any feed-forward graph of the ensemble $g(m, n)$ other than $\tilde{\mathcal{G}}_n(m)$ is obtained by removing edges of $\tilde{\mathcal{G}}_n(m)$. This edge removal process will necessarily result in a reduction of uncertainty.$\square$

Lemma 2: *The entropy of $\tilde{\mathcal{G}}_n(m)$ is bounded as:*

$$H(\tilde{\mathcal{G}}_n(m)) \leq \log m + \frac{\log 2}{2}(n - m - 1).$$

Proof. This Lemma is a direct consequence of the application of Jensen's inequality -see equation (4). Given the labeling proposed above, and starting from eq. (9) the number of paths in $\tilde{\mathcal{G}}_n(m)$ from M to $v_i \in V \setminus M$ will be:

$$
\begin{aligned}
|\Pi(v_i)| &= \sum_{j \leq L(\mathcal{G})} \sum_{l : v_l \in M} \left[\mathbf{A}^T(\tilde{\mathcal{G}}_n(m)) \right]_{il}^j \\
&= \sum_{l : v_l \in M} \sum_{j \leq i} \binom{i}{j} \\
&= m \sum_{j \leq i} \binom{i}{j} = m \cdot 2^{i-1}.
\end{aligned}
$$

Using Jensen's inequality, one has that:

$$
\begin{aligned}
h(v_i) &\leq \log |\Pi(v_i)| \\
&= \log m + (i - 1) \log 2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
H(\tilde{\mathcal{G}}_n(m)) &= \frac{1}{n - m} \sum_{v_i \in V \setminus M} h(v_i) \\
&\leq \frac{1}{n - m} \Big((n - m) \log m + \\
&\quad + \frac{(n - m)(n - m - 1)}{2} \log 2 \Big) \\
&= \log m + \frac{\log 2}{2}(n - m - 1).\square
\end{aligned}
$$

2. The explicit form of entropies of $\tilde{\mathcal{G}}_n(m)$

The next step is to derive the mathematical expression corresponding to $\tilde{\mathcal{G}}_n(m)$. We begin with a lemma which describes the structure of $\Phi(\tilde{\mathcal{G}}_n(m))$ -see section III B, equation (10).

Lemma 3: *Given the labeling proposed in the section (A 1) the probabilities defining the matrix $\Phi(\tilde{\mathcal{G}}_n(m))$ are given by:*

$$\phi_{ik} = \frac{1}{m + k} \quad (k < i).$$

Proof. The first observation is that, for any node $v_i \in V \setminus M$ the probability to reach one maximal is

$$\frac{1}{m}.$$

What about v_1, i.e., the first node we find after the maximal set? We observe that, from the viewpoint of node v_i $(i > 1)$, the situation is completely analogous to the configuration where there are $m + 1$ maximal nodes, since, due to the property of acyclicity, the probability to pass through v_1 does not depend on what happens *above* v_1. Therefore:

$$\phi_{i1} = \frac{1}{m + 1}.$$

Running the reasoning from v_1 to v_{i-1}, we find that:

$$\phi_{ik} = \frac{1}{m + k} \quad (k < i).\square$$

Interestingly, for $k < i$, ϕ_{ik} is invariant, no matter the value of i. This leads matrix $\Phi(\tilde{\mathcal{G}}_n(m))$ to be:

$$
\Phi(\tilde{\mathcal{G}}_n(m)) = \begin{pmatrix}
1 & 0 & 0 & \dots & 0 \\
\frac{1}{m+1} & 1 & 0 & \dots & 0 \\
\frac{1}{m+1} & \frac{1}{m+2} & 1 & \dots & 0 \\
\dots & & & & \dots \\
\dots & & & & \dots \\
\frac{1}{m+1} & \frac{1}{m+2} & \frac{1}{m+3} & \dots & 1
\end{pmatrix}, \quad \text{(A1)}
$$

Now we are ready to obtain the final expression of the entropy of $\tilde{\mathcal{G}}_n(m)$. First, we observe that

$$h_{loc}(v_k) = \log(m + k - 1).$$

Therefore, inserting it and (A1) into eq. (14), we obtain, after some algebra:

$$H(\tilde{\mathcal{G}}_n(m)) = \frac{n}{n - m} \sum_{i \leq n - m} f(v_i),$$

where $f(v_i)$ is a function $f : V \setminus M \to \mathbb{R}^+$,

$$f(v_i) = \frac{\log(m + i - 1)}{m + i}. \quad \text{(A2)}$$

We can see that the value entropy is reduced to the computation of the average of f over the set $V \setminus M$. Having $\tilde{\mathcal{G}}_n(m)$ n nodes, from which m of them are maximal ones, we will refer to this average as $\langle f_n(m) \rangle$, defined as:

$$\langle f_n(m) \rangle = \frac{1}{n - m} \sum_{i \leq n - m} f(v_i) \quad \text{(A3)}$$

3. Absolute maxima of entropies

Theorem 1: *Let*

$$\tilde{\mathcal{G}}_n(1), ..., \tilde{\mathcal{G}}_n(n-1)$$

be the set whose members are the maximum entropic structures of ensembles of DAGs having n nodes, $g(1,n), ..., g(n-1,n)$, respectively. Then, $\forall i \neq 2$

$$H(\tilde{\mathcal{G}}_n(2)) > H(\tilde{\mathcal{G}}_n(i))$$

for values of $n \gg 1$.

Proof. Let us first notice that:

$$Q(\tilde{\mathcal{G}}_n(2)) = Q(\tilde{\mathcal{G}}_n(1)),$$

enabling us to derive the first inequality:

$$\begin{aligned} H(\tilde{\mathcal{G}}_n(2)) &= \frac{1}{n-2}Q(\tilde{\mathcal{G}}_n(1)) \\ &> \frac{1}{n-1}Q(\tilde{\mathcal{G}}_n(1)) \\ &= H(\tilde{\mathcal{G}}_n(1)). \end{aligned}$$

Once we demonstrated that $H(\tilde{\mathcal{G}}_n(2)) > H(\tilde{\mathcal{G}}_n(1))$, we proceed to demonstrate that $H(\tilde{\mathcal{G}}_n(2)) > H(\tilde{\mathcal{G}}_m(3))$. To this end, let us first observe a key property of f, defined in eq. (A2). Indeed, we observe that $(\forall \epsilon > 0)(\exists k_\epsilon) : (\forall k > k_\epsilon)$,

$$f(v_k) < \epsilon,$$

provided that n is large enough. From this property, and since $\langle f_n(m) \rangle$ is an average -see eq. (A3)- we can be sure

that $(\exists n^*) : (\forall n > n^*)$,

$$\frac{\log 2}{3} > \langle f_n(3) \rangle,$$

by choosing appropriately n in such a way that we have enough terms lower than a given ϵ to obtain the above desired result. Thus, from eq. (A3) and knowing that

$$H(\tilde{\mathcal{G}}_n(2)) - H(\tilde{\mathcal{G}}_n(3)) \propto \frac{\log 2}{3} - \langle f_n(3) \rangle,$$

(with proportionality factor equal to $n/(n-2)$) we can conclude that

$$H(\tilde{\mathcal{G}}_n(2)) > H(\tilde{\mathcal{G}}_n(3)).$$

The general case easily derives from the same reasoning, since:

$$H(\tilde{\mathcal{G}}_n(k)) - H(\tilde{\mathcal{G}}_n(k+1)) \propto \frac{\log(k+1)}{k} - \langle f_n(k+1) \rangle,$$

and thus, we can conclude that:

$$(\forall k \leq 2) \ \ H(\tilde{\mathcal{G}}_n(k)) > H(\tilde{\mathcal{G}}_n(k+1)).$$

This closes the demonstration that, for n large enough, $\tilde{\mathcal{G}}_n(2)$ is the most entropic graph. According to numerical computations, this demonstration holds provided that $n > 14$. $\square$

Acknowledgments

---

[1] G. Csardi, K. J. Strandburg, L. Zalanyi, J. Tobochnik, and P. Erdi, Physica A **374**, 783 (2007).

[2] B. Karrer and M. E. J. Newman, Phys Rev Lett **102**, 128701 (2009).

[3] S. Lehmann, B. Lautrup, and A. D. Jackson, Phys. Rev. E **68**, 026113 (2003).

[4] S. Valverde, R. V. Solé, M. A. Bedau, and N. Packard, Phys Rev E Stat Nonlin Soft Matter Phys **76**, 056118 (2007).

[5] R. Clay, *Nonlinear networks and systems* (John Wiley & Sons Inc, New York, 1971).

[6] S. Haykin, *Neural Networks : a Comprehensive Foundation* (Prentice-Hall. London, 1999).

[7] H. Frank and I. T. Frisch, *Communication, transmission and transportation networks* (Addison-Wesley (Reading Mass), 1971).

[8] I. Rodríguez-Iturbe and A. Rinaldo, *Fractal River Basins. Chance and Self-organization* (Cambridge University Press. Cambridge., 1997).

[9] D. Bonchev and D. H. Rouvray, *Complexity in Chemistry, Biology, and Ecology* (Springer, New York., 2005).

[10] C. H. Bennett, IBM J. Res. Dev. **17**, 525 (1973).

[11] R. Landauer, IBM Journal of Research and Development **5**, 183 (1961).

[12] W. Fontana and L. W. Buss, Proc Natl Acad Sci U S A **91**, 757 (1994).

[13] S. J. Gould, *Wonderful Life: The Burgess Shale and the Nature of History* (W. W. Norton & Company. New York, 1990).

[14] S. Gavrilets, *Fitness Landscapes and the Origin of Species (MPB-41) (Monographs in Population Biology)* (Princeton University Press, 2004).

[15] S. Kauffman and S. Levin, J Theor Biol **128**, 11 (1987).

[16] R. V. Sole, P. Fernandez, and S. A. Kauffman, International Journal of Developmental Biology **2003**, 685 (2003).

[17] S. R. de Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (North-Holland. Amsterdam, 1962).

[18] G. Lebon, D. Jou, and J. Casas-Vázquez, *Understanding Nonequilibrium Thermodynamics* (Springer, Berlin, 2008, 2008).

[19] P. Schuster, Complexity **In press** (2010).

[20] A. Shreim, P. Grassberger, W. Nadler, B. Samuelsson, J. E. S. Socolar, and M. Paczuski, Phys. Rev. Lett. **98**, 198701 (2007).

[21] B. A. Huberman and T. Hogg, Phys. D **2**, 376 (1986).

[22] R. Badii and A. Politi, Physical Review Letters (1997).

[23] K. Anand and G. Bianconi, Phys. Rev. E **80**, 045102 (2009).

[24] M. Dehmer, Appl. Artif. Intell. **22**, 684 (2008).

[25] M. Dehmer, S. Borgert, and F. Emmert-Streib, PLoS ONE **3**, e3079 (2008).

[26] E. Estrada, Phys. Rev. E **80**, 026104 (2009).

[27] E. Schneidman, S. Still, M. J. Berry, and W. Bialek, Phys. Rev. Lett. **91**, 238701 (2003).

[28] R. V. Solé and S. Valverde, in *Networks: Structure, Dynamics and Function, Lecture Notes in Physics.* (Springer-Verlag, 2004), pp. 189–210.

[29] R. E. Ulanowicz, *Growth and Development: Ecosystems Phenomenology.* (Springer, New York., 1986).

[30] B. Bollobás, *Modern Graph Theory* (Springer, 1998), corrected ed.

[31] J. Gross and J. Yellen, *Graph Theory and its applications* (CRC, Boca Raton, Florida, 1998).

[32] J. Kelley, *General Topology*, Graduate Texts in Mathematics, 27, 1975 (Van Nostrand, 1955).

[33] P. Suppes, *Axiomatic Set Theory* (Dover. New York, 1960).

[34] R. B. Ash, *Information Theory* (New York. Dover, 1990).

[35] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley and Sons. New York, 1991).

[36] A. I. Khinchin, *Mathematical Foundations of Information Theory* (Dover, New York, 1957).

[37] C. E. Shannon, Bell System Technical Journal **27**, 379 (1948).

[38] M. C. Lagomarsino, P. Jona, and B. Bassetti, in *CMSB* (2006), pp. 227–241.

[39] C. Rodríguez-Caso, B. Corominas-Murtra, and R. V. Solé, Mol Biosyst **5**, 1617 (2009).

[40] F. V. Jensen, *Bayesian Networks and Decision Graphs*, Information Science and Statistics (Springer, 2002).

[41] D. J. Futuyma, *Evolution* (Sinauer Associates. Sunderland, 2005).

[42] T. Dagan, Y. Artzy-Randrup, and W. Martin, Proc Natl Acad Sci U S A **105**, 10039 (2008).

[43] T. Dagan and W. Martin, Philos Trans R Soc Lond B Biol Sci **364**, 2187 (2009).

[44] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry, Third Edition (North-Holland Personal Library)* (North Holland, 2007), 3rd ed.

[45] It is important to notice that the results reported in this paper are independent on the number of connected components displayed by the DAG. However, we will tacitly assume that one single connected component linking all nodes is present, unless the contrary is indicated. An intuitive statement guides our choice: Two unconnected components are causally independent and, therefore, they must be treated as independent entities.

[46] We observe that matrix $\mathbf{B}^T$ is the matrix corresponding to a Markov process [44] depicting a random walker walking against the flow

[47] Throughout this appendix we will refer simply as *entropy* to the entropy derived in equation (14).