INTERFACE

Research

Author for correspondence:
Sara Imari Walker
e-mail: sara.i.walker@asu.edu

THE ROYAL SOCIETY
PUBLISHING

New scaling relation for information transfer in biological networks

Hyunju Kim[1], Paul Davies[1] and Sara Imari Walker[1,2,3,4]

[1]BEYOND: Center for Fundamental Concepts in Science, [2]School of Earth and Space Exploration, and [3]ASU-SFI Center for Biosocial Complex Systems, Arizona State University, Tempe, AZ, USA
[4]Blue Marble Space Institute of Science, Seattle, WA, USA

We quantify characteristics of the informational architecture of two representative biological networks: the Boolean network model for the cell-cycle regulatory network of the fission yeast *Schizosaccharomyces pombe* (Davidich *et al.* 2008 *PLoS ONE* **3**, e1672 (doi:10.1371/journal.pone.0001672)) and that of the budding yeast *Saccharomyces cerevisiae* (Li *et al.* 2004 *Proc. Natl Acad. Sci. USA* **101**, 4781–4786 (doi:10.1073/pnas.0305937101)). We compare our results for these biological networks with the same analysis performed on ensembles of two different types of random networks: Erdös–Rényi and scale-free. We show that both biological networks share features in common that are not shared by either random network ensemble. In particular, the biological networks in our study process more information than the random networks on average. Both biological networks also exhibit a scaling relation in information transferred between nodes that distinguishes them from random, where the biological networks stand out as distinct even when compared with random networks that share important topological properties, such as degree distribution, with the biological network. We show that the most biologically distinct regime of this scaling relation is associated with a subset of control nodes that regulate the dynamics and function of each respective biological network. Information processing in biological networks is therefore interpreted as an *emergent property of topology* (*causal structure*) and *dynamics* (*function*). Our results demonstrate *quantitatively* how the informational architecture of biologically evolved networks can distinguish them from other classes of network architecture that do not share the same informational properties.

1. Introduction

Living systems are often described in terms of logic modules, information flows and computation [1]. Such informational language is used in fields as diverse as evolutionary biology [2], neuroscience [3], embryogenesis [4], colonial decision-making in eusocial insects [5] and protein–protein interaction networks [6], to name just a few. There are two perspectives on the ontological status of information in biology. The first regards information as *intrinsic* to the operation of living systems [7]. If this stronger viewpoint proves correct, then life would necessarily be classified as distinct from other kinds of physical systems, as we know of no other class of physical system where information is necessary to specify its state [8].[1] There is increasingly support for the strong viewpoint [12–17], and it is well known that information must of course be instantiated in physical degrees of freedom: 'information is physical!' as Landauer [18,19] famously proclaimed. However, there also remains much support for a second, weak viewpoint, which takes the perspective that while information is certainly a useful metaphor to describe biological systems, ultimately, all of biological complexity is, at least in principle, fully reducible to known physics [20].

A necessary step to resolve the debate over the status of information in living processes is to investigate if, and if so how, information operates in complex biological systems. In particular, a convincing case for the strong viewpoint should *quantitatively* satisfy two conditions:

(1) biological systems must be demonstrated to somehow be unique in their informational architecture, when compared with other classes of physical systems and

(2) information must be shown necessary to the execution of biological function—that is to say, information must be shown to matter to matter.

The necessity of the first condition is perhaps obvious: if information is fundamental to biological organization, then a strong signal should appear when contrasting information in living systems with the same concept of information as applied to non-living systems. The challenge is to define the appropriate concept of 'information' relevant to biological organization. The second criterion is less immediately obvious, but is clarified by a simple example. An informational pattern, such as the sequence of bases necessary to specify the GNRA tetraloop 'GAGA', is readily copyable to other states of matter, e.g. such as this page of text. This kind of pattern cannot therefore be unique to life in the sense of the strong view, as other states of matter can share the same informational pattern. This may be one reason why attempts to quantify biological complexity in terms of Shannon entropy have been relatively unsuccessful [21] (for example, it is well known that genome size, which can be correlated with Shannon information content, does not readily map to organismal complexity [22]). What, if anything, distinguishes life from other complex physical systems, as quantified by 'information', must therefore necessarily tie information to doing something, e.g. *to the causal structure underlying function* [7,23–26]. It is for this reason that we use the terminology 'informational architecture' rather than use 'information' or 'informational pattern' herein, as architecture implies physical structure, whereas mere patterns are not necessarily tied to their physical instantiation.

In what follows, we focus on two well-studied Boolean network models for real biological systems: the network regulating the cell cycle of the fission yeast (*Schizosaccharomyces pombe*) [27] and that of the budding yeast (*Saccharomyces cerevisiae*) [28]. Our approach is to compare the informational architecture of these Boolean network models with ensembles of random networks to identify informational properties that might distinguish the biological networks from random. To quantify characteristics of the informational architecture of the biological and random networks, we use information-theoretic analyses, detailing how information is processed in the execution of biological function. Our approach aims, in part, to address a research programme laid out in a seminal paper by Nurse [1], wherein he calls for more focused understanding on *flows of information* within biological networks. Therefore, we focus on the concept of 'information flows', interpreted as *information processing* and measured by Schreiber's transfer entropy (TE) [29], as a relevant concept of information that allows us to address conditions (1) and (2) above. We note that our approach is different from those attempting to quantify 'functional information' [23,30], the nature of which has been widely debated in its own right, which is not addressed here. Instead, we explicitly focus on *how information is processed in the execution of function*. There are many other informational measures in the literature based on Shannon entropy or Kolmogorov–Chaitin complexity, which are beyond the scope of this study, that could be tested to determine whether they can distinguish the informational structure of biological systems from random in a similar manner to the analysis presented here (see [31] for an analysis by us using other information measures). We also note that

while there has been a great deal of emphasis placed on understanding the 'logic of life', quantitative results have thus far been primarily topological *or* dynamical. Examples include analysis of the topology of network motifs—or logic circuits—necessary to biological function [32,33], or dynamical network features such as the robustness of the global attractor landscape [34]. The approach we present herein is distinct from these previous efforts in that we explicitly address information processed, which, as we will show, should be viewed as an emergent property of networks that arises from the integration of topology *and* dynamics.

For both the *S. pombe* and *S. cerevisiae* cell-cycle networks, we calculate the information transferred between pairs of nodes in the execution of function (defined in the Boolean network models as the *dynamics* of each network), and contrast the results with the same analysis performed on ensembles of random networks drawn from two different classes: Erdös–Rényi (ER) networks and scale-free (SF) networks. The ER random networks share no structural bias with the cell-cycle networks other than the size of the networks. The SF random network ensembles, by contrast, maintain the same number of activation and inhibition links for each individual node as the cell-cycle networks. The SF networks therefore share important topological properties with the biological networks, including degree distribution (the rank ordering of the number of edges per node), often cited as a distinguishing feature of many real-world networks including biological networks [35–37]. In our analysis, features shared by the biological network and SF networks, but not the ER networks, can be concluded to arise as a result of topological features, whereas those observed in the biological network that are not shared with the SF networks should be regarded as arising specifically owing to network features distinctive to biological function (presumably generated via natural selection). By comparing results for the biological network with both ER and SF random networks, we are therefore able to distinguish which features of the informational architecture of biological networks arise solely as a result of network topology (e.g. degree distribution, which is shared with the SF networks but not the ER networks) and which are specific to a biological function.

We find that both cell-cycle networks share commonalities in their informational architecture that set them far apart from their random network counterparts. Both cell-cycle networks are outliers with respect to the total information processed: the biological networks process more information than the vast majority of random networks in either ensemble, suggestive of evolutionary optimization for information processing. The most striking feature uncovered in our analysis is a scaling relation for the distribution of information transfer between nodes, which for the cell-cycle networks is statistically different from that observed for either ensemble of random networks. This is despite the topological resemblance between the cell-cycle networks and SF networks in our study. We identified the regime where the biological networks differ most significantly from the random networks in their scaling relation, and characterized the patterns in information transfer relative to the function (dynamics) of each network. Our results show that the biologically most distinctive regime is dominated by information transfer through a subset of control nodes, called the control kernel, that are associated with regulation of the function of each biological network. We also investigated how the causal structure (topology) of the cell-cycle networks affects information transfer when compared with the random
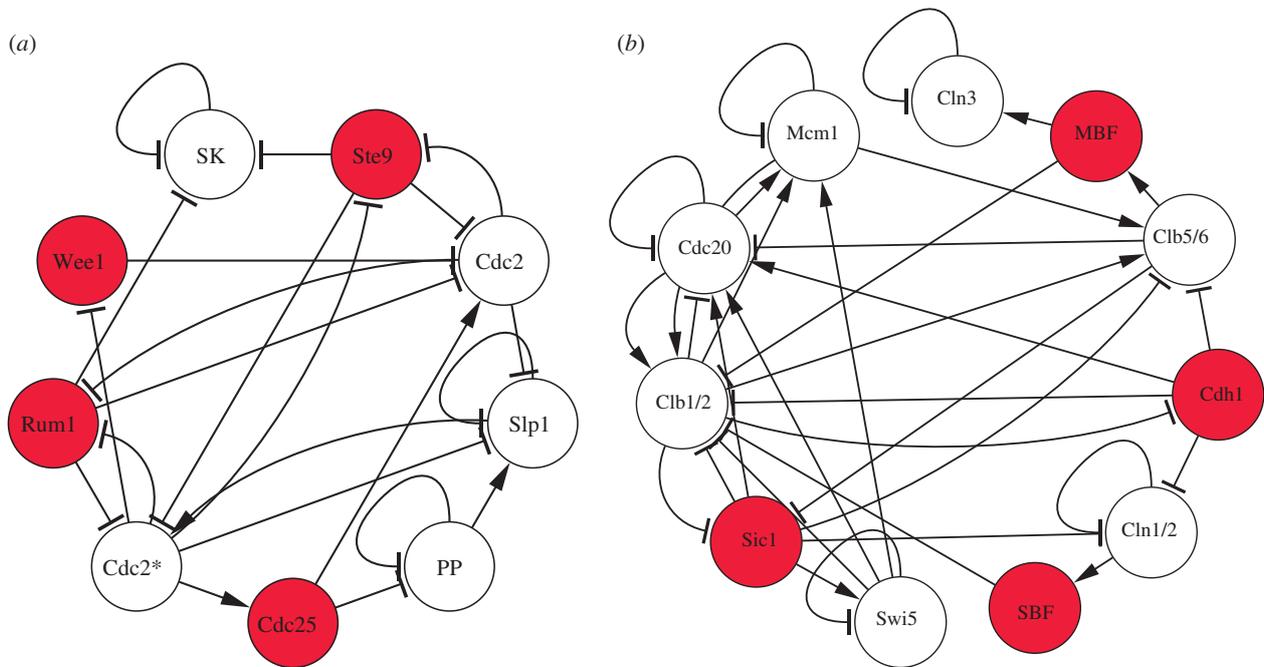
Figure 1. Boolean network models of (*a*) fission and (*b*) budding yeast cell-cycle regulation. Nodes represent the regulatory proteins, and edges denote two types of biochemical interactions between nodes: activation (ended with an arrow) and inhibition (ended with a bar). The nodes coloured red are the *control kernel*, which regulates the global behaviour of each network when pinned to specific values. Figure adapted from [27,28,45].

networks and found that the SF structure, shared by the cell-cycle and SF networks, uses long-range correlations more than direct causal interactions between nodes for information processing, unlike what is observed for ER networks.

Our results are suggestive of previously unidentified information-based organizational principles that go beyond topological considerations, which may be critical to biological function. The uncovered informational architecture for the biological networks therefore potentially offers an operational means to quantify life in terms of its informational properties. The results presented thus open a new *quantitative* framework in support of the strong viewpoint on the status of *information in biology*, that complements other approaches [12–15,26,38,39], by definitively demonstrating how the informational architecture of biologically evolved networks can distinguish them from other classes of network architecture that do not share the same informational properties.

2. Model and methods

The study presented herein is a first attempt, to the best of our knowledge, to identify patterns in information processing that might be distinctive of biological organization, when compared with other classes of physical systems, and in turn to connect these patterns to causal structure (topology) and function (dynamics). Our analysis requires an integrative synthesis of a number of distinct areas, including Boolean network models of biological function, information-theoretic analysis for distributed computation, sampling networks with topological constraints and control of cellular behaviour. We briefly describe each herein.

2.1. Boolean network models for the cell-cycle regulatory process

Boolean network models have proven, in many cases, to provide accurate models for biological function [40–42]. They are also the most readily tractable network models for information-theoretic analysis, because each node may take on only one of two discrete states—'0' or '1' [43,44]. They are thus ideal for our study. In this study, we focus on the Boolean network models for cell-cycle regulation in the fission yeast *S. pombe* [27] and the budding yeast *S. cerevisiae* [28]. Regulation of cellular division is a central aspect of cellular function. During the cellular division process, the cell passes through a number of phases, G1–S–G2–M, which collectively constitute the cell cycle dictating the steps of cellular division to produce two daughter cells from a single parent cell. During the G1 stage, the cell grows, and if conditions are favourable, then it can commit to division. During the S stage, DNA is replicated. In the G2 stage, there is a 'gap' between DNA replication (in the S phase) and mitosis (in the M phase) where the cell continues to grow. In the M stage, the cell undergoes mitosis, and two daughter cells are produced. After the M stage, the daughter cells enter G1 again, thereby completing the cycle.

The Boolean network models for the regulation of the cell-cycle process in the fission yeast *S. pombe* and the budding yeast *S. cerevisiae* are shown in figure 1. Each node corresponds to one protein among the small subset of key regulatory proteins involved in each respective cell cycle. The state of node i, $S_i(t) \in \{0, 1\}$, indicates whether the given protein is present ('1') or not ('0'). Biochemical causal interactions between proteins are denoted by edges in the network. Time advances through the cycle via synchronous updating, with the states of all nodes updated at discrete timesteps according to the following rule:

$$S_i(t+1) = \begin{cases} 1, & \sum_j a_{ij} S_j(t) > \theta_i \\ 0, & \sum_j a_{ij} S_j(t) < \theta_i \\ S_i(t), & \sum_j a_{ij} S_j(t) = \theta_i, \end{cases} \quad (2.1)$$

where a_{ij} is the edge weight between node i and j ($a_{ij} = -1$ for inhibition links and $a_{ij} = 1$ for activation links), and θ_i is the threshold for node i. The threshold for all nodes in both

Table 1. Constraints for constructing two different classes of random networks that retain features of the causal structure of a reference cell-cycle network.

	Erdös – Rényi (ER) networks	scale-free (SF) networks
size of network (total number of nodes, inhibition and activation links)	same as the cell-cycle network	same as the cell-cycle network
nodes with a self-loop	same as the cell-cycle network	same as the cell-cycle network
threshold for each node	same as the cell-cycle network	same as the cell-cycle network
number of activation and inhibition links for each node	*not* the same as the cell-cycle network (→no structural bias)	same as the cell-cycle network (→SF structure)

networks is 0, with the exception of Cdc2/13 and Cdc2/13* in the fission yeast network, which have thresholds of −0.5 and 0.5, respectively.

Although numerous Boolean network models for biological systems exist, we specifically focus on these two networks because they are small and accurately model an essential biological function. The small network size, with approximately 10 nodes each, permits statistically reliable comparison of results for the biological networks with the average properties of randomized networks of the same size (number of nodes and edges), owing to the relatively small ensemble size of comparable random networks. Thus, for these two networks, we can readily address condition (1) as posed above, by making a meaningful comparison between each biological network and an ensemble of random networks with similar size and topological features. Connecting information processing and function, as laid out by criterion (2), is also tenable for both models. For both cell-cycle networks, there is a direct connection between the dynamics on these networks and the corresponding biological function: both networks correctly reproduce the sequences of protein states corresponding to the phases of the respective cell cycle (see [27,28] for details). Therefore, any distinctive patterns uncovered in our analysis of informational architecture can be related to dynamics, and consequently, the biological function of each network.

We note, in particular, that the task of addressing condition (2) is, in general, not a trivial one. Our approach is to connect information processing to topology *and* dynamics, that is to the causal mechanisms of each biological network that define its function (as we will show in addressing condition (1), topology alone is not sufficient to quantify information processing in the biological networks). Stated differently, condition (2) requires identifying how information transfer might depend on causal structure in non-trivial ways such that *information processing is intrinsic to function* (and therefore that information processing matters to the world of matter). We note that the edges connecting nodes for both cell-cycle networks are modelled based on experimental data detecting a direct causal interaction between the proteins they represent [27,28]. Comparing information transfer between nodes connected by an edge to that between nodes that are not directly interacting therefore provides us insights into how correlations are distributed relative to causal structure in the execution of function.

Our analysis uncovers distinct patterns in information processing associated with a few key nodes of each cell-cycle network, called the *control kernel*, which regulate the dynamics of each respective cell-cycle network [46]. When the values of the control kernel are fixed to their values in the primary attractor associated with biological function, the entire network converges to that attractor, regardless of the initial state of the network—that is to say, the control kernel nodes regulate the function (dynamics) of the network. The control kernel thus provides us a local mechanism for regulating the global behaviour of the network. The control kernel nodes are highlighted in red in figure 1 for each cell-cycle network. Control kernels have been found in a number of biological networks by Kim *et al.* [45], suggestive that they may be a generic feature of Boolean models for regulatory networks. To address condition (2), we show how information transfer is related to the causal mechanisms of both cell cycles by determining the distribution of information transfer among pairs of nodes with a causal connection (edge) and without, and more specifically, we show that the features that most distinguish the biological from random networks are associated with information transfer through the control kernel nodes.

We note that despite the fact that budding and fission yeasts are closely related genetically [46], closely related genes between the two organisms can play vastly different functional roles [47]. In addition, while the two networks share similar overall dynamics in terms of the dominating largest attractor [27,28], they show significant differences in their underlying biochemical machinery [48]. Thus, for the purposes of our study, we view them as two independent examples of biological networks with related function. Studying both networks in parallel provides a comparative analysis to look for features common to biological networks that are not shared by their random counterparts.

2.2. Construction of random Boolean networks with biological constraints

To identify any features distinctive to biologically evolved networks, we compare the results of our information-theoretic analysis on both cell-cycle networks with the same analysis performed on ensembles of two types of random networks: ER networks and SF networks. For meaningful comparison, both classes retain certain features of the biological network's causal structure as summarized in table 1. Both ensembles of random networks share the same number of nodes and the same total number of activation and inhibition edges as the biological network. The ER networks are fully randomized with respect to the number of activation and inhibition links per node—that is, they have no structural bias (e.g. no hubs). The terminology of 'ER network' used here is meant to indicate that these networks are generated by using an ER random graph model, where every pair of nodes is connected by an edge with a fixed probability [49,50]. We note that our ER networks are equivalent to the class of networks commonly

referred to as 'random networks' in other literature. However, we use the term 'ER networks' to distinguish them from another class of randomly sampled network used in this study—SF networks. By contrast to the ER networks, the SF networks in our study share the same number of activation and inhibition links as the cell-cycle network *for each node*, and therefore share the same degree distribution, defined as the rank ordering of the number of (in-directed and out-directed) edges per node [51]. We generate the SF networks by using an edge-swapping algorithm applied to the cell-cycle networks to preserve the degree sequence of each cell-cycle network in the sampled SF networks (for a more general method to generate networks with a given degree sequence, see [52–54]).

In our study, the term 'SF network', unlike its common definition, does not mean that sample networks in the SF ensemble exhibit power-law degree distributions. Owing to their small size, even the degree distributions of the biological cell-cycle networks do not follow a power law. Instead, the term 'SF' as used in this paper emphasizes that the sample networks have the *same exact degree sequence* as each cell-cycle network (table 1). Because having the same degree sequence is a sufficient condition for sharing a degree distribution, for larger biological networks that are indeed truly SF, the analogous random graphs would therefore also be scale-free.

Previously, many features of biological networks (e.g. robustness to random failure, small number of attractors, etc.) [35] have been explained as arising from their SF structure. Here, we are in particular interested in going beyond an SF structure and addressing whether information processing (which we attribute to topology and dynamics) is a more distinctive feature of biological networks than global topology (SF structure) alone. For this reason, we purposefully do not use a fully randomized network ensemble having no structural features in common with the biological network in our comparison and use the same update rule in equation (2.1) for all networks in our study. The comparison of the cell-cycle networks with both ensembles of ER and SF random networks allows us to isolate the contribution attributable to topological features such as degree distribution, which the biological networks share with the SF networks but not the ER networks in our sample, from any informational structure that may arise solely owing to network architecture which is peculiar to biological function. A unique feature of our analysis is that each random network in our ER and SF ensembles is constructed relative to a specific biological network. We are therefore able to identify characteristics of the biological networks that arise owing to the intrinsic complexity emerging from the patterns in connectivity among nodes in the cell-cycle networks that distinguish them from similarly constructed random graphs, rather than addressing extrinsic properties such as network size, which are beyond the scope of this study. However, important future work may involve investigations of the dynamical process for non-living systems, even with different extrinsic properties, for comparison with biological systems in terms of our analysis of informational architecture.

2.3. Quantifying information processing

Our information-theoretic analysis focuses on quantifying 'information processing' in the biological and ER and SF random networks. Our motivation is to capture Nurse's [1]

notion of 'information flows'[2] as a concept of information relevant to biological function. We adopt the notion of information processing as implemented in information dynamics, a formalism for quantifying the component operations of computation in dynamical systems, using the tools of information theory. In information dynamics, information processing is quantified using Schreiber's TE [29,43,56,57], a directional measure of information transfer. TE from a *source* node Y to a *destination* node X is defined as the reduction in uncertainty owing to knowledge of state of Y about the future state of X, above the reduction in uncertainty provided by knowledge about the past states of X. TE from Y to X can be written as the mutual information between y_n and x_{n+1} conditioned on k previous states of X

$$T_{Y \to X}(k) = \sum_{(x_n^{(k)}, x_{n+1}, y_n) \in \mathcal{A}_0} p(x_n^{(k)}, x_{n+1}, y_n)$$
$$\times \log_2 \frac{p(x_{n+1}|x_n^{(k)}, y_n)}{p(x_{n+1}|x_n^{(k)})}, \qquad (2.2)$$

where $\mathcal{A}_0$ indicates the set of all possible patterns of sets of states $(x_n^{(k)}, x_{n+1}, y_n)$ and $x_n^{(k)}$ denotes $(x_n, \ldots, x_{n-k+1})$, the vector of k previous states of destination X at timestep $n+1$. In addition, y_n and x_{n+1} represent the state of Y at the current timestep and the state of X at the next timestep, respectively. Directionality arises owing to the asymmetry between timesteps for the state of the source and the destination in equation (2.2). Owing to this asymmetry, TE can be used to measure 'information flows'. It is therefore more appropriate for our purposes of quantifying distinct features of information processing in biological networks than related measures such as mutual information, which measure correlations only, without reference to directionality. We note that the causal structure for the networks studied herein is fully described by their edges as noted in §2.1. TE is correlational and does not necessarily reflect direct causal interactions between pairs of nodes, but can take on non-zero values even for pairs of nodes that do not directly interact [56], which as we show below is important to distinguishing the biological from random networks. Calculation of TE is not level specific, and thus the approach presented here is expected to generalize to most, if not all, biological networks.

The probability distributions in equation (2.2) are defined as the relative frequency of each pattern of states over the timeseries of dynamical states of the network. In our study, we obtained the probability distributions for each cell-cycle network and the sampled networks from the ER and SF random network ensembles necessary to calculate the TE between every ordered pair of nodes (i, j), given by $T_{i \to j}$ for history lengths $k = 1, 2, \ldots, 8$. To do so, we generated every possible trajectory for each network by applying the updating rule in equation (2.1) up through 20 timesteps for all the 2^n possible initial network states (i.e. all possible combinations of binary states for nodes in the network, where n is the total number of nodes in the particular network under study) and counted the frequency of each pattern of states in equation (2.2) (see [39] for an explicit example of calculating TE based on frequency distributions). We chose 20 timesteps as the length of the trajectory generated from each initial state, because it is sufficiently long to capture transient dynamics for networks before converging on a fixed point (an attractor) for the cell cycle and for the vast majority of random networks.

3. Results

3.1. Scaling relation for information transfer in yeast cell-cycle networks

To reiterate, we use TE as a candidate measure for identifying features potentially distinctive to biological networks, focusing our analysis on the concept of 'information flows' that may be particular to biological function as suggested by Nurse [1]. For every ordered pair of nodes (i, j) in both cell-cycle networks, we calculated TE from node i to node j, $T_{i \to j}$ (as described in §2.3) and ranked the pair (i, j) according to its measured value of TE. The same analysis was performed on each of the 1000 sampled networks in the ensembles of ER and SF random networks as a point for comparison to identify any features peculiar to the biological networks.

The resulting scaling relations are shown in figure 2, where biological networks are highlighted in red for fission yeast (top) and budding yeast (bottom). The ensemble averages for the scaling relations of ER and SF random networks generated with reference to each cell cycle are shown in green and blue, respectively, in each respective panel. Error bars represent the standard deviation over the ensembles of random networks (averages are over 1000 sampled networks). For the results shown in figure 2, the history lengths $k = 2$ and $k = 5$ were selected because these history lengths show the most distinctive scaling distribution when compared with the two random network ensembles for the fission and budding yeast networks, respectively (see the electronic supplementary material for scaling relations over different history lengths). The scaling distributions reveal a nonlinear relationship between the information transferred between pairs of nodes (y-axis) and their relative rank (x-axis), for each of the network classes studied—biological, SF and ER. The scaling relation is most striking for the biological networks (red), which are significant outliers with respect to either of the ER or SF ensembles.

Figure 2 shows that ER networks have much less information transfer than SF or biological networks. This result suggests that SF structure—as is the case for the SF and biological networks, but not the ER networks—plays a significant role in increasing information transfer within a network. The deviation between ER networks and SF networks or the cell-cycle networks may be expected owing to the differences in topological features, which the TE is sensitive to. However, surprisingly, SF structure alone does not account for the high level of information transfer between nodes in the biological networks. The biological networks differ from the SF networks despite their common topological features. With the exception of the few highest ranked node pairs, the biological networks exhibit information transfer that is several standard deviations larger than the corresponding rank in the SF ensemble for the majority of ranks with TE $\neq 0$. The excess TE observed in the biological networks in figure 2 deviates between 1σ and 5σ from that of the SF networks, with a trend of increasing divergence from the SF ensemble for lower-ranked node pairs that still exhibit correlations (e.g. where TE > 0). We define χ as the set of node pairs whose TE deviates > 2σ from the SF networks, indexed by rank. The ranks that deviate more than 2σ for the fission yeast are $\chi_f = \{9, 10 \ldots 30\}$ and for budding yeast they are $\chi_b = \{31, 32 \ldots 68\}$ (highlighted between the dashed lines in each panel in figure 2). Patterns in the
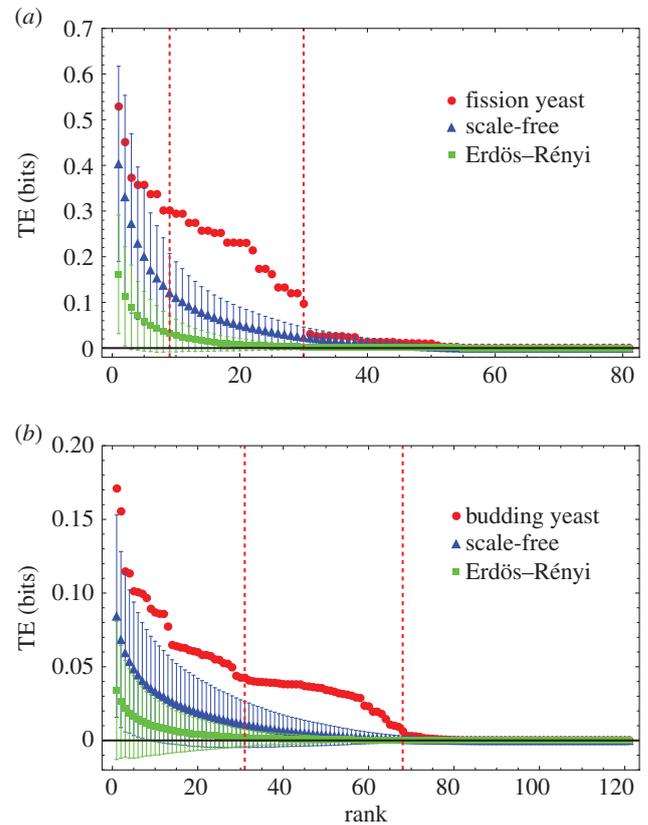

Figure 2. Scaling of information processing (TE) among pairs of nodes for cell-cycle (red), ER (green) and SF (blue) networks. History lengths for computing TE were $k = 2$ and $k = 5$ for the fission and the budding yeast, respectively. The averages and the standard deviation for each of the random network ensembles are computed for a sample of 1000 networks. Regions between dashed lines denote χ for each cell-cycle network. (a) Fission yeast. (b) Budding yeast.

biological distribution also exhibit plateaus, suggestive that there exist subgroups wherein informational flow is evenly distributed, as opposed to a few dominating informational connections. We analyse this substructure in §3.4 in terms of the dynamics (function) of the cell-cycle networks.

3.2. Total information processed

We also calculated the total information processed by the cell-cycle networks and compared with that of individual instances of sampled networks from the ER and SF random network ensembles. We define a quantity called *total information processed*, denoted by I_p, which is the sum of the TE between all pairs of nodes (i, j) for an individual network, $I_p = \sum_{(i,j)} T_{i \to j}$, for a given history length k ($k = 2$ and $k = 5$ for the fission and the budding yeast, respectively). Accordingly, the total information processed by the fission and budding yeast cell-cycle networks is $I_p = 8.09$ and $I_p = 3.51$, respectively—shown as red lines in figure 3. The frequency distributions of networks associated with I_p for the two sets of SF and ER networks are shown in figure 3. Comparing the two distributions, it is evident that, on average, SF networks process more information than ER networks, with higher frequencies for networks with larger values of I_p. Figure 3 also shows that the biological networks process more information than most random networks in the ensembles. More specifically, the fission yeast cell-cycle network lies outside of 95% of the SF networks and 100% of the ER
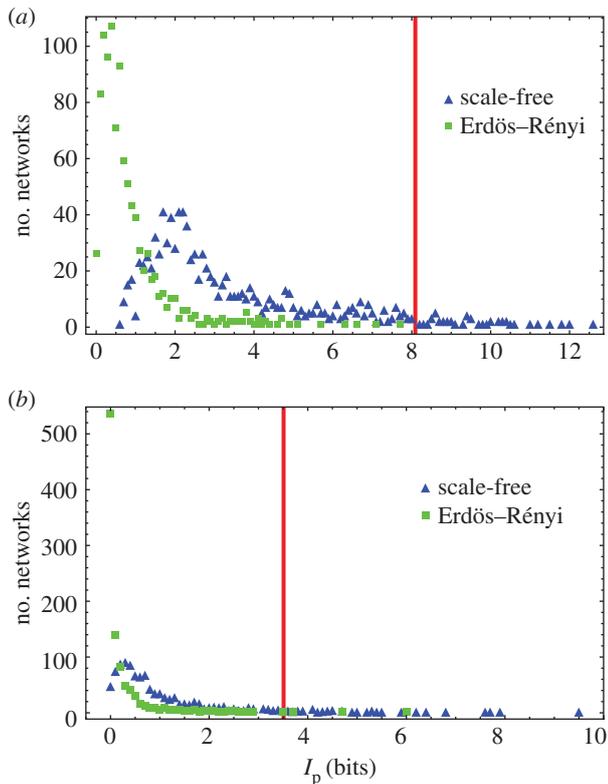
Figure 3. Distributions for total information processed (sum of TE for all pairs of nodes) for the ensembles of ER (green) and SF (blue) networks. Each data point represents the number (*y*-axis) of individual networks within the respective ensemble with a given amount of total information transferred (*x*-axis). The red line indicates the total information processed for the (*a*) fission and (*b*) budding yeast cell-cycle network.

Figure 4. Classification of all pairs of nodes within the cell-cycle (red), SF (blue) and ER (green) networks by correlation (TE > 0 or TE = 0) and causal inter-action (edge or no edge). Each data bar indicates the percentage of node pairs (on *y*-axis) within each category. (*a*) Fission yeast. (*b*) Budding yeast.

networks. Similarly, the budding yeast cell-cycle network is in the 95.1% of SF networks and the 99.5% of ER networks. This result indicates that the cell-cycle networks are highly optimized, with only 1/2000 ER networks and 1/200 SF networks displaying comparable total information processed in a random draw.

3.3. Distribution of information processing over causal structure

The correlations measured with TE can arise owing to direct causal effect (e.g. via an edge) or statistical correlations between two nodes that are not directly connected via a causal interaction (e.g. via long-range correlations embedded in the distributed causal structure of a network). We note that because we are using TE in our analysis, the measured corre-lations are across both *space* (between nodes) and *time* (between discrete timesteps)—they therefore differ from cor-relations typically associated with critical phenomena as applied to networks or other physical systems, which are usually strictly spatial (e.g. such as mutual information or in the study of physical phase transitions, correlation length).

To identify whether the distinctive features of the cell-cycle networks shown in figures 2 and 3 arise owing to infor-mation transfer along edges, or longer-range correlations, we classified each pair of nodes as either having a direct causal interaction (connected by an edge) or not (no edge). We also classified them as being correlated (TE > 0) or not (TE = 0). The results of this classification scheme are shown in table 2, and are very similar for both cell-cycle networks.
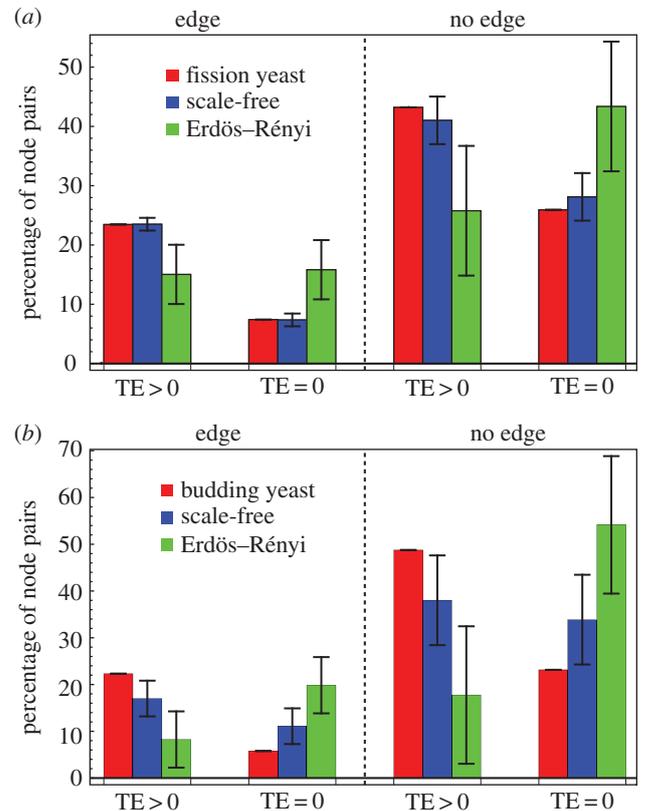
Table 2. The distribution of TE within the cell-cycle networks and corresponding SF and ER networks, classified by pairs of nodes that are correlated (TE > 0) or not (TE = 0) and causally interacting (edge) or not (no edge). The values indicate the ratio of the number of node pairs in each category to the total number of node pairs for each cell-cycle network.

	edge (%)	no edge (%)
(*a*) fission yeast		
TE > 0	23.46	43.21
TE = 0	7.41	25.93
(*b*) budding yeast		
TE > 0	22.31	48.76
TE = 0	5.79	23.14

In both cases, the majority (more than 40%) of node pairs are correlated via information transfer, even though they are not causally connected by an edge. Roughly 25% of node pairs have a causal interaction with information trans-fer, or no interaction and no information transfer. The remaining minority of nodes exhibit a causal interaction with no corresponding transfer of information (causation without correlation). The same analysis was applied to the sampled networks from the ER and SF random network ensembles constructed relative to each cell cycle; the averages over 1000 networks in each ensemble are shown in figure 4

Table 3. Possible information flows between sets of nodes classified as control kernel (CK; highlighted in red in figure 1) or non-control kernel (NCK).
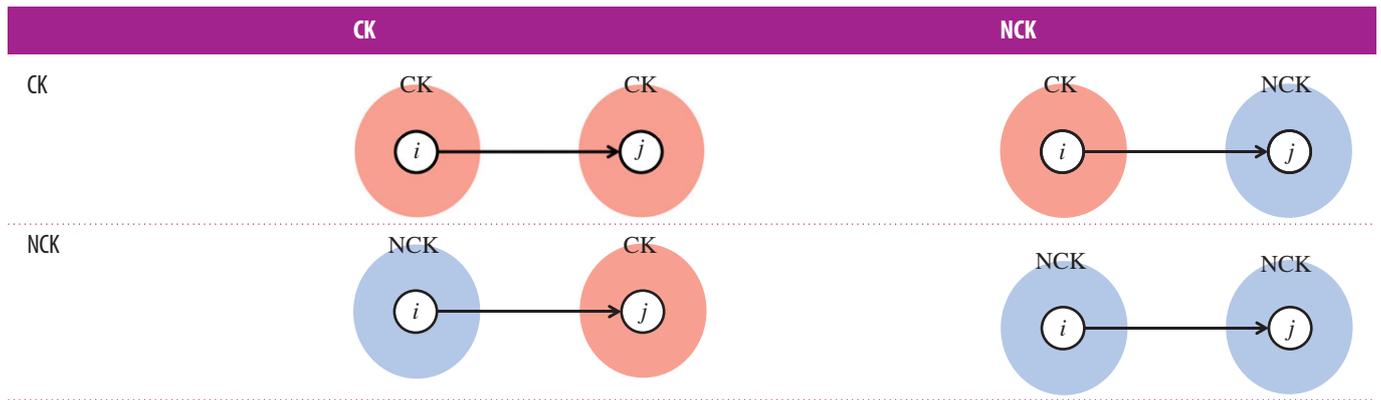
(see the electronic supplementary material for detailed data). Figure 4 shows a distinctive transition in the distribution of correlations moving from the ER to SF network ensembles. In the ER networks, the majority of node pairs that are not connected by an edge are also not correlated. By contrast, the majority of node pairs in SF networks are correlated even though they are not directly connected, and this pattern is even more prominent for the biological cell-cycle networks. Biological networks therefore appear to be highly optimized for long-range correlations among node pairs.

3.4. Information transfer and regulation of biological function

We analysed the scaling relations for the cell-cycle networks in figure 2, with particular focus on the biologically distinct regime χ, in terms of the global causal structure of each biological network. While the local causal structure of these networks is fully articulated by the edges shown in figure 1, we view the global causal structure as the trajectories of network states associated with the attractor landscape of each network. In their role regulating the dynamics of each respective cell-cycle network—and thus regulating function—the control kernel nodes mediate the connection between the local and global causal structure of each network (see also discussion in [31]).

Here, we are interested in how the control kernel nodes—as drivers of global dynamics and functional regulators—play a role in *information transfer* within the cell-cycle networks. To determine the role of a control kernel in information processing by the biological networks, we divided all nodes in each cell-cycle network into two groups: CK and NCK, where CK denotes the set of control kernel nodes identified in [45] and NCK is the complement of CK, i.e. non-control kernel nodes. Therefore, all ordered pairs of nodes (i, j) in each network fall into one of four groups shown in table 3 depending on whether i or j belongs to CK or NCK. We specified the groups for each node pair in the scaling patterns for information transfer for both biological networks (red in figure 2), as shown in figure 5.

The highest ranked regimes of the cell-cycle scaling relations shown in figure 2, where the biological networks *differ least* from the SF networks, are dominated by information transfer between NCK nodes (NCK → NCK, shown in aqua in figure 5). The most biologically distinctive regime (χ_f and χ_b) is, by contrast, dominated by information
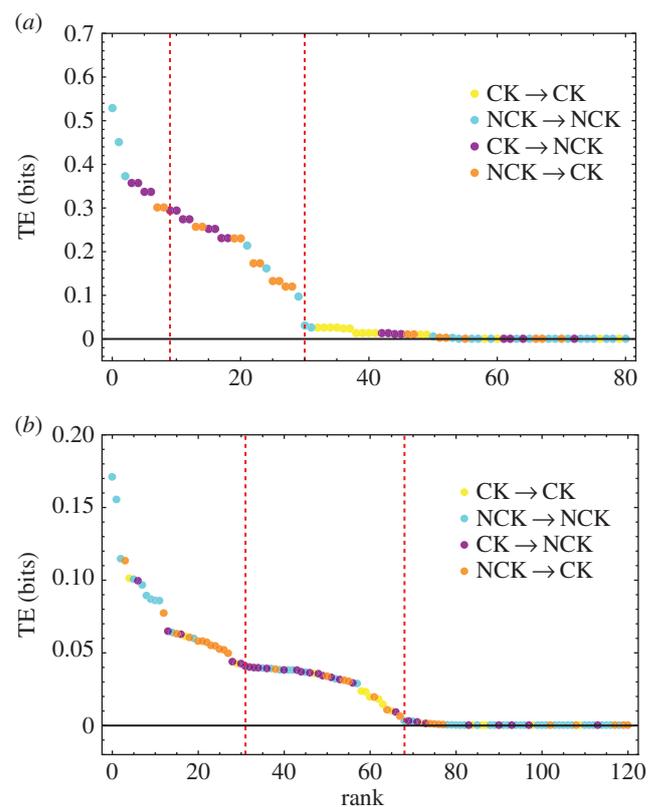

Figure 5. Scaling relations for information transfer for the fission yeast (*a*) and budding yeast (*b*) cell-cycle networks. Data shown are the same as in figure 2 with data points divided into four classes of information transfer: CK → CK (yellow), CK → NCK (purple), NCK → CK (orange) and NCK → NCK (aqua).

transfer between CK nodes and NCK nodes, i.e. CK → NCK (purple) and NCK → CK (orange; see electronic supplementary, table S1). *The scaling regime that deviates most from the SF networks in the biological networks is dominated by information transfer through the control kernel nodes.* This suggests that for both cell-cycle networks, the patterns in information processing observed to be most distinct to the biological networks are strongly affected by the presence of the control kernel, and hence must be associated with the regulation of function.

4. Conclusion

Support for the viewpoint that information is not merely descriptive, but instead a key *physical* aspect of the operation

of living systems (strong view) requires that at least two conditions are satisfied: (i) biological systems must be demonstrated to somehow be unique in their informational architecture, when compared with other classes of physical systems and (ii) 'information' must be shown to be intrinsic to the operation of biological systems, e.g. to the execution of biological function. Our analyses presented here provide us first quantitative data addressing these two conditions, with results lending support to the strong view that biological systems are distinctive in their informational architecture, where the relevant concept of 'information' is information processing as quantified by Schreiber's TE [29].

Our results indicate that SF structure—characterizing the biological and SF networks in our study, but not the ER networks—plays a significant role in increasing information transfer within a network. This result is particularly interesting when considered within the context of the widespread observations of SF structure in various biological networks [35–37]. The ubiquity of SF networks is *prima facie* explained as a result of their robustness properties. However, the high levels of information transfer in SF networks when compared with ER networks uncovered in this study indicate an alternative explanatory hypothesis may also hold true—that is, that *SF networks are widespread in biological systems as a result of their amplified information processing*.

Significantly, the enhancement of SF networks over ER networks is not sufficient to explain the information transfer observed for the biological networks. Condition (1) is therefore satisfied by the statistically significant difference between the distribution of 'information flows' for biological networks and that of ensembles of ER and SF networks: SF structure taken alone is not sufficient to explain the distribution of information processing for either biological network. This result has important implications for modelling of biological function, which has thus far primarily focused on topological *or* dynamic properties. In particular, topological features of SF networks—such as power-law degree distribution—are often viewed as sufficient to capture essential features of biological organization due precisely to the fact that they are observed in a number of disparate biological phenomena ranging from metabolic networks to the Internet [35,37]. Our results indicate that an additional criterion for accurately modelling biological function—network models should include the *dynamics of information transfer* in their construction, and in particular that models should optimize information transfer between nodes, because this feature distinguishes the biological networks in our study from generic SF networks. Our analysis shows that this optimization arises owing to the presence of a subset of nodes that can 'control' the dynamics of the network on its attractor landscape, which is a key feature of the distinctive features of biological informational architecture reported here. The newly discovered informational scaling relation is an emergent property of networks that arises from the integration of *topology* and *dynamics* as a hallmark of localized control, which cannot be accounted for solely by either topology or dynamics alone.

Additionally, the biological networks in our study are outliers in terms of the total information processed, as quantified by I_p, in each network in our study. This concept of information (information processing) satisfies the condition of being not readily copyable to other physical systems, because it is the physical instantiation (causal mechanisms)

that give rise to the observed patterns in information flows, and thus is a candidate for satisfying condition (2). Our analysis also, in part, directly addresses condition (2), by connecting the observed distinctive features of the informational architecture of biological networks in figure 2 to their causal structure and their function. The enhancement of information transfer between nodes for biological networks, when compared with SF or ER, is primarily owing to correlations between nodes which are *not* directly causally connected (non-local)—a frequent signature of collective, or critical states of organization. Typically, criticality is described in terms of long-range correlations in space. Herein, we have shown that for the cell-cycle networks, correlations are in *space* and *time*, and are associated with information processing. Interestingly, both biological networks studied have very similar patterns for the distribution of TE among causally connected node pairs (table 2). It is an open question whether this pattern is indicative of cell-cycle function, or a more general pattern of biological organization that might be characteristic of networks with other functions. In particular, because the cell cycle is optimized for *processing information through time* as a mechanism for keeping track of phases during cellular division, a question of interest is whether other networks, optimized for *processing spatial information* (for example, the genetic regulation of embryo development [58]), exhibit the same or different informational patterns in their distribution of correlations among causal edges. Future work will determine whether these patterns are ubiquitous features of biological organization as well as connect the patterns uncovered herein to more traditional approaches to physics by considering the connections to energy efficiency and computation as performed by the biological networks to generate testable predictions regarding the physics underlying living matter [59–61].

An important feature of both biological networks in this study, which is shared with other regulatory networks, is the control kernel, defined as a subset of nodes that regulate the attractor dynamics associated with biological function. Our analysis has revealed that most of the biologically distinct regimes of the information transfer scaling relations, χ_f and χ_b for the fission and budding yeast, respectively, is attributable to the presence of the control kernel. Information transfer in or out of the control kernel dominates the biologically distinct regime for both networks. Furthermore, for the budding yeast network, the ranks deviating most from random are primarily those corresponding to information transfer between control kernel nodes. Although not conclusive that information is intrinsic to function (e.g. that condition (2) holds), our results clearly indicate that the patterns in information processing unique to the biological networks in our study are attributable to the regulation of function by a few key nodes. Interestingly, these nodes have other properties consistent with their information-theoretic interpretation: in particular, the control kernel provides us a mechanism for *distinguishability* among attractor states for the biological networks. As noted by Kim *et al.* [45], the set of control kernel nodes takes on a unique and distinct state in every attractor state in the networks studied. They thus provide a means for coarse-graining the state space in a functionally relevant manner (such that the primary attractor associated with function is distinguishable from other attractor states). That the network organizes information flows through these nodes as it executes dynamics of the cell cycle is a highly non-trivial

feature of both networks' informational structure. We also note that the control kernel nodes were effectively discovered by a causal intervention in the manner of Pearl [62], i.e. by fixing the subset of nodes to their value in the biologically relevant attractor. Combined with our results, this suggests interesting connections between information transfer and top-down causal regulation [63] of biological function, discussed in more depth by us in [31].

We hypothesize that the features reported herein may be common to biological networks of different function, and in particular, that scaling relations in information transfer may be a universal hallmark of biological organization characteristic of life not only here on the Earth, but also potentially elsewhere. Our results are suggestive of previously unidentified information-based organizational principles that go beyond topological considerations such as SF structure, and may be critical to biological function. They thus open a new *quantitative* framework in support of the strong view that information is indeed intrinsic to life [1,2,10,24], by demonstrating how the informational architecture of biologically evolved networks can distinguish them from other complex physical systems that do not exhibit the same informational properties. This suggests that new 'informational laws', with biological informational architecture at their core, might be necessary to account for life and perhaps even the emergence of life.

Endnotes

[1]With the exception of the trivial sense in which 'information' is used to describe every physical system. From the perspective of the strong viewpoint, information is not only descriptive, but also explanatory [9–11].

[2]Herein, we use the terms 'flow', 'transfer' and 'processing' interchangeably in reference to information, and quantify this concept of information using Schreiber's transfer entropy [29]. This is more informal language than the technical meaning of 'information flow' as formulated by Ay & Polani [55], which is a measure of causal flows. Because we do not implement the Ay and Polani measure, herein, we use 'flow' to directly connect our quantitative results with the notion of 'information flow' as introduced by Nurse (which does not necessarily imply direct causal interaction).

References

1. Nurse P. 2008 Life, logic and information. *Nature* **454**, 424–426. (doi:10.1038/454424a)

2. Goldenfeld N, Woese C. 2010 Life is physics: evolution as a collective phenomenon far from equilibrium. (http://arxiv.org/abs/1011.4125)

3. Lizier JT, Heinzle J, Horstmann A, Haynes J-D, Prokopenko M. 2011 Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity. *J. Comput. Neurosci.* **30**, 85–107. (doi:10.1007/s10827-010-0271-2)

4. Levin M. 2014 Molecular bioelectricity: how endogenous voltage potentials control cell behavior and instruct pattern regulation *in vivo*. *Mol. Biol. Cell* **25**, 3835–3850. (doi:10.1091/mbc.E13-12-0708)

5. Waters JS, Fewell JH. 2012 Information processing in social insect networks. *PLoS ONE* **7**, e40337. (doi:10.1371/journal.pone.0040337)

6. Bray D. 1995 Protein molecules as computational elements in living cells. *Nature* **376**, 307–312. (doi:10.1038/376307a0)

7. Ellis GFR, Noble D, O'Connor T. 2012 Top-down causation: an integrating theme within and across the sciences? *Interface Focus* **2**, 1–3. (doi:10.1098/rsfs.2011.0110)

8. Walker SI, Davies P. 2013 The algorithmic origins of life. *J. R. Soc. Interface* **6**, 20120869. (doi:10.1098/rsif.2012.0869)

9. Deutsch D. 2013 Constructor theory. *Synthese* **190**, 4331–4359. (doi:10.1007/s11229-013-0279-z)

10. Deutsch D, Marletto C. 2015 Constructor theory of information. *Proc. R. Soc. A* **471**, 20140540. (doi:10.1098/rspa.2014.0540)

11. Marletto C. 2015 Constructor theory of life. *J. R. Soc. Interface* **12**, 20141226. (doi:10.1098/rsif.2014.1226)

12. Danchin A. 2009 Bacteria as computers making computers. *FEMS Microbiol. Rev.* **33**, 3–26. (doi:10.1111/j.1574-6976.2008.00137.x)

13. Yockey HP, Platzman RL, Quastler H. 1958 *Symposium on information theory in biology*. New York, NY: Pergamon Press.

14. Antoine D. 2002 *The Delphic boat: what genomes tell us*. Cambridge, MA: Harvard University Press.

15. Binder PM, Danchin A. 2011 Life's demons: information and order in biology: what subcellular machines gather and process the information necessary to sustain life? *EMBO Rep.* **12**, 495–499. (doi:10.1038/embor.2011.83)

16. Toyabe S, Sagawa T, Ueda M, Muneyuki E, Sano M. 2010 Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality. *Nat. Phys.* **6**, 988–992. (doi:10.1038/nphys1821)

17. Vedral V. 2010 *Decoding reality: the universe as quantum information*. Oxford, UK: Oxford University Press.

18. Landauer R. 1991 Information is physical. *Phys. Today*, **44**, 23. (doi:10.1063/1.881299)

19. Landauer R. 1996 The physical nature of information. *Phys. Lett. A* **217**, 188–193. (doi:10.1016/0375-9601(96)00453-7)

20. Dawkins R. 1986 *The blind watchmaker: why the evidence of evolution reveals a universe without design*. New York, NY: WW Norton & Company.

21. Danchin A. 1996 On genomes and cosmologies. In *Integrative approaches to molecular biology* (eds J Collado-Vides, B Magasanik, TF Smith), pp. 91–111. Cambridge, MA: MIT Press.

22. Eddy SR. 2012 The c-value paradox, junk DNA and encode. *Curr. Biol.* **22**, R898–R899. (doi:10.1016/j.cub.2012.10.002)

23. Griffiths PE, Pocheville A, Calcott B, Stotz K, Kim H, Knight R. 2015 Measuring causal specificity. *Philos. Sci.* **82**, 529–555. (doi:10.1086/682914)

24. Auletta G, Ellis GFR, Jaeger L. 2008 Top-down causation by information control: from a philosophical problem to a scientific research programme. *J. R. Soc. Interface* **5**, 1159–1172. (doi:10.1098/rsif.2008.0018)

25. Davies PCW. 2012 The epigenome and top-down causation. *Interface Focus* **2**, 42–48. (doi:10.1098/rsfs.2011.0070)

26. Roederer J. 2006 *Information and its role in nature*. Berlin, Germany: Springer Science & Business Media.

27. Davidich MI, Bornholdt S. 2008 Boolean network model predicts cell cycle sequence of fission yeast. *PLoS ONE* **3**, e1672. (doi:10.1371/journal.pone.0001672)

28. Li F, Long T, Lu Y, Ouyang Q, Tang C. 2004 The yeast cell-cycle network is robustly designed. *Proc. Natl Acad. Sci. USA* **101**, 4781–4786. (doi:10.1073/pnas.0305937101)

29. Schreiber T. 2000 Measuring information transfer. *Phys. Rev. Lett.* **85**, 461–464. (doi:10.1103/PhysRevLett.85.461)

30. Hazen RM, Griffin PL, Carothers JM, Szostak JW. 2007 Functional information and the emergence of biocomplexity. *Proc. Natl Acad. Sci. USA* **104**(Suppl. 1), 8574–8581. (doi:10.1073/pnas.0701744104)

31. Walker SI, Kim H, Davies PCW. In press. The informational architecture of the cell. *Phil. Trans. R. Soc. A.*

32. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. 2002 Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827. (doi:10.1126/science.298.5594.824)

33. Alon U. 2006 *An introduction to systems biology: design principles of biological circuits*. Boca Raton, FL: CRC Press.

34. Huang S, Eichler G, Bar-Yam Y, Ingber DE. 2005 Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* **94**, 128701. (doi:10.1103/PhysRevLett.94.128701)

35. Barabási A-L, Albert R. 2002 Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97. (doi:10.1103/RevModPhys.74.47)

36. Eguiluz VM, Chialvo DR, Cecchi GA, Baliki M, Apkarian AV. 2005 Scale-free brain functional networks. *Phys. Rev. Lett.* **94**, 018102. (doi:10.1103/PhysRevLett.94.018102)

37. Albert R. 2005 Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957. (doi:10.1242/jcs.02714)

38. Walker SI, Davies P, Ellis GFR (eds). 2016 *From matter to life: information and causality*. Cambridge, UK: Cambridge University Press.

39. Davies P, Walker SI. Submitted. The hidden simplicity of biology.

40. Albert R, Othmer HG. 2003 The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J. Theor. Biol.* **223**, 1–18. (doi:10.1016/S0022-5193(03)00035-3)

41. Espinosa-Soto C, Padilla-Longoria P, Alvarez-Buylla ER. 2004 A gene regulatory network model for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. *Plant Cell* **16**, 2923–2939. (doi:10.1105/tpc.104.021725)

42. Mendoza L, Thieffry D, Alvarez-Buylla ER. 1999 Genetic control of flower morphogenesis in *Arabidopsis thaliana*: a logical analysis. *Bioinformatics* **15**, 593–606. (doi:10.1093/bioinformatics/15.7.593)

43. Lizier JT, Prokopenko M, Zomaya AY. 2008 The information dynamics of phase transitions in random Boolean networks. In *Proc. 11th Int. Conf. on the Synthesis and Simulation of Living Systems*, pp. 374–381. Cambridge, MA: MIT Press.

44. Lizier JT, Pritam S, Prokopenko M. 2011 Information dynamics in small-world Boolean networks. *Artif. Life* **17**, 293–314. (doi:10.1162/artl_a_00040)

45. Kim J, Park S-M, Cho K-H. 2013 Discovery of a kernel for controlling biomolecular regulatory networks. *Sci. Rep.* **3**, 2223. (doi:10.1038/srep02223)

46. Tyson JT, Chen K, Novak B. 2001 Network dynamics and cell physiology. *Nat. Rev. Mol. Cell Biol.* **2**, 908–916. (doi:10.1038/35103078)

47. Forsburg SL. 1999 The best yeast? *Trends Genet.* **15**, 340–344. (doi:10.1016/S0168-9525(99)01798-9)

48. Simanis V. 2003 Events at the end of mitosis in the budding and fission yeasts. *J. Cell Sci.* **116**, 4263–4275. (doi:10.1242/jcs.00807)

49. Erdös P, Rényi A. 1959 On random graphs. I. *Publ. Math. Debrecen* **6**, 290–297.

50. Erdös P, Rényi A. 1960 On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17–61.

51. Maslov S, Sneppen K. 2002 Specificity and stability in topology of protein networks. *Science* **296**, 910–913. (doi:10.1126/science.1065103)

52. Kim H, Toroczkai Z, Erdős PL, Miklós I, Székely LA. 2009 Degree-based graph construction. *J. Phys. A, Math. Theor.* **42**, 392001. (doi:10.1088/1751-8113/42/39/392001)

53. Del Genio CI, Kim H, Toroczkai Z, Bassler KE. 2010 Efficient and exact sampling of simple graphs with given arbitrary degree sequence. *PLoS ONE* **5**, e10012. (doi:10.1371/journal.pone.0010012)

54. Kim H, Del Genio CI, Bassler KE, Toroczkai Z. 2012 Constructing and sampling directed graphs with given degree sequences. *New J. Phys.* **14**, 023012. (doi:10.1088/1367-2630/14/2/023012)

55. Ay N, Polani D. 2008 Information flows in causal networks. *Adv. Complex Syst.* **11**, 17–41. (doi:10.1142/S0219525908001465)

56. Lizier JT, Prokopenko M. 2010 Differentiating information transfer and causal effect. *Eur. Phys. J. B* **73**, 605–615. (doi:10.1140/epjb/e2010-00034-5)

57. Wang XR, Miller JM, Lizier JT, Prokopenko M, Rossi LF. 2011 Measuring information storage and transfer in swarms. In *Proc. 11th European Conf. on the Synthesis and Simulation of Living Systems*, pp. 838–845. Cambridge, MA: MIT Press.

58. Yuh C-H, Bolouri H, Davidson EH. 1998 Genomic *cis*-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* **279**, 1896–1902. (doi:10.1126/science.279.5358.1896)

59. Still S, Sivak DA, Bell J, Crooks GE. 2012 Thermodynamics of prediction. *Phys. Rev. Lett.* **109**, 120604. (doi:10.1103/PhysRevLett.109.120604)

60. Prokopenko M, Lizier JT, Price DC. 2013 On thermodynamic interpretation of transfer entropy. *Entropy* **15**, 524–543. (doi:10.3390/e15020524)

61. Prokopenko M, Lizier JT. 2014 Transfer entropy and transient limits of computation. *Sci. Rep.* **4**, 5394. (doi:10.1038/srep05394)

62. Pearl J. 2000 *Causality: models, reasoning and inference*, vol. 29. Cambridge, MA: MIT Press.

63. Walker SI. 2014 Top down causation and the rise of information in the emergence of life. *Information* **5**, 424–439. (doi:10.3390/info5030424)