Information storage, loop motifs, and clustered structure in complex networks

Joseph T. Lizier,[1,2,*] Fatihcan M. Atay,[1,†] and Jürgen Jost[1,3,‡]

[1]*Max Planck Institute for Mathematics in the Sciences, Inselstraße 22, 04103 Leipzig, Germany*

[2]*CSIRO Information and Communications Technology Centre, P. O. Box 76, Epping, New South Wales 1710, Australia*

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

(Received 29 November 2011; revised manuscript received 21 February 2012; published 15 August 2012)

We use a standard discrete-time linear Gaussian model to analyze the information storage capability of individual nodes in complex networks, given the network structure and link weights. In particular, we investigate the role of two- and three-node motifs in contributing to local information storage. We show analytically that directed feedback and feedforward loop motifs are the dominant contributors to information storage capability, with their weighted motif counts locally positively correlated to storage capability. We also reveal the direct local relationship between clustering coefficient(s) and information storage. These results explain the dynamical importance of clustered structure and offer an explanation for the prevalence of these motifs in biological and artificial networks.

DOI: 10.1103/PhysRevE.86.026110 PACS number(s): 89.75.Fb, 89.70.Cf, 87.19.lo, 02.10.Ox

I. INTRODUCTION

Recognizing that network structure gives rise to dynamics, but dynamics represents the specific action of a network, much recent work has focused on studying the *complexity* of dynamics on various network structures; see, e.g., [1–3]. Such investigations can reveal the dynamic capabilities of well-known structures, e.g. small-world networks [2,4,5]. We take an information-theoretic approach to this issue, exploring *computational properties* (e.g., information storage and transfer [6–8]) as a function of network structure. Such an approach is highly appropriate [5], since these terms are well defined and can be measured on any type of network time-series data, and are also meaningful and well understood (e.g., information transfer as directed coupling between two nodes), especially in comparison to general notions of complexity, and also because computation is the language in which dynamics on networks is often described (e.g., claims that small-world structures have "maximum capability to store, process, and transfer information" [9]).

Here we focus on information storage, which is a key operation in intrinsic distributed computation, underpinning periodic behavior, stability, and the concept of memory [6]. It is considered an important aspect of the dynamics of many natural and man-made processes on complex networks, e.g., in human brain networks [10,11] and artificial neural networks [12,13]. Yet a general understanding of the relationship between network structure and information storage remains absent.

We seek to provide such an understanding, focusing on the role of motif structures (small subnetwork configurations) [14,15] in facilitating information storage. Intuitively, one would expect directed cycles (or feedback loop motifs) to play a pivotal role, since *isolated* directed cycles clearly provide paths for nodes to store information in neighbors and retrieve it at a later time. There is much circumstantial evidence for such intuition to extend to the role of such loops *embedded* in networks, e.g., in supporting long cyclic patterns in cellular automata [6]; underpinning better performance of artificial recurrent neural networks than feedforward networks on tasks requiring memory (e.g., by echo state networks [12,13] and Elman networks [16]); and in the strong correlation between clustering in underlying topology and information storage capability in a study of small-world Boolean networks [5]. Interestingly, Song *et al.* [17] find a greater than expected likelihood for reciprocal connectivity and connected three-neuron motifs in a mammalian cortex; Sporns [11] states that this may be due to an inherent preference for local connectivity, but also points out that different motifs could support "different modes of information processing" and that clustered nodes (qualitatively) seem to share information and are likely to promote "a functionality coherent brain system."

Despite such intuition, there is no *direct* analytic evidence for the role of loop motifs in information storage when embedded in networks, nor for which if any other motifs are involved. Here, we will follow Barnett *et al.* [1,2] in making analytic inference of information-theoretical measures of dynamics on networks. This involves linear stationary Gaussian processes, which are a simplification as compared to real-world processes, but can be viewed as approximating the weakly coupled near-linear regime, and are commonly used in neuroscience for large-scale modeling [1].

We measure information storage by individual nodes within a network, and analytically reveal the roles of network motifs up to size 3. We prove that information storage is dominated by directed cycles (as per intuition) and feedforward loop motifs involving the given node, being directly proportional to these local weighted motif counts. We also show the direct local relationship between the clustering coefficient and storage. These results help explain the prevalence of these motifs in the biological and artificial networks described above.

II. INFORMATION-THEORETIC MEASURES ON LINEAR GAUSSIAN NETWORKS

The (differential) entropy for a multivariate normal distribution $\mathbf{X}$ (of N variables) is, e.g., [18]: $H(\mathbf{X}) = \frac{1}{2}\ln[(2\pi e)^N|\Omega|]$,

*lizier@mis.mpg.de
†fatay@mis.mpg.de
‡jjost@mis.mpg.de

where $|\Omega|$ is the determinant of the $N \times N$ covariance matrix $\Omega = \overline{\mathbf{X}(n)^T \mathbf{X}(n)}$, and the overbar "represents an average over the statistical ensemble" [1] at times n. Any standard information-theoretic measure of the variables (at the same time step) can then be obtained from sums and differences of these joint entropies. Furthermore, information-theoretic measures relating variables over a time difference s can be formed from the *lagged covariance matrix* $\Omega(s)$.

Extending original attempts by Tononi *et al.* [3], Barnett *et al.* [1] have shown how to obtain these covariance matrices given a network topology and connection strengths on each link, and assuming linear relationships between nodes driven by Gaussian noise (which approximates "the statistical structure of signals sampled from the environment" [1]). In this paper we will consider *discrete-time processes*, in particular the multivariate Gaussian autoregressive process on a network of N nodes [1]:

$$\mathbf{X}(n+1) = \mathbf{X}(n) \times C + \mathbf{R}(n), \tag{1}$$

at time steps n. Here, $C = [C_{ji}]$ is the $N \times N$ connectivity matrix (or *weighted* adjacency matrix), where C_{ji} is the weight of the directed connection from node j to i, and I is the $N \times N$ identity matrix. The current node value $\mathbf{X}(n)$ is a row vector here. $\mathbf{R}(n)$ is uncorrelated mean-zero unit-variance Gaussian noise. Barnett *et al.* [1] show that the covariance matrix Ω satisfies $\Omega = C^T \Omega C + I$, with the solution then in general obtained from the power series expansion (insofar as it converges)

$$\Omega = I + C^T C + (C^2)^T C^2 + \cdots = \sum_{u=0}^{\infty} (C^u)^T C^u. \tag{2}$$

It is important that the system is stationary only when we have $|\lambda| < 1$ for all eigenvalues λ of C [i.e., a spectral radius $\rho(C) < 1$]. This is also sufficient condition for the convergence of Eq. (2), following the similar argument of Barnett *et al.* for the counterpart of Eq. (2) in continuous-time systems in [1] (i.e., using Gelfand's theorem with any consistent matrix norm).

We contribute that for a lag s (an integer in the discrete-time system) the lagged covariance matrix is

$$\Omega(s) = \overline{\mathbf{X}(n)^T \mathbf{X}(n+s)} = \Omega C^s, \tag{3}$$

noting that covariances with the noise terms disappear under averaging, and that $\Omega(0) = \Omega$.

III. COMPUTING INFORMATION STORAGE

We will measure the *active information storage* $A(X_i)$ [6,8] at each node i in the network. $A(X)$ is defined for variable X as the average mutual information $\langle i(x_n^{(k)}; x_{n+1}) \rangle$ between its joint past k states $x_n^{(k)} = \{x_{n-k+1}, \ldots, x_{n-1}, x_n\}$ and its next state x_{n+1} at time step $n+1$ (as $k \to \infty$) [6,8]:

$$A(X) = \lim_{k \to \infty} \sum_{x_n^{(k)}, x_{n+1}} p\left(x_n^{(k)}, x_{n+1}\right) \log_2 \frac{p\left(x_n^{(k)}, x_{n+1}\right)}{p\left(x_n^{(k)}\right) p(x_{n+1})}.$$

In comparison to the use of eigenvalue-based decay rates (e.g., see [19]) to *infer* memory, $A(X_i)$ measures storage at each node rather than providing a network-wide measure, is a direct measure of information rather than an inference, is model-free and may be applied to any (nonlinear) time series, and (as we shall see) can be analytically related to local network motifs. We will reconsider this comparison in the Discussion. Further, $A(X)$ is a subcomponent of the *excess entropy* E [20], a known complexity measure. While $E(X)$ captures the *total* information from the past that is used in the future, $A(X)$ captures the stored information *in use* in the next state, and so is directly comparable to other dynamic quantities of computation (e.g., information transfer) [6–8].

We will compute $A(X)$ using its formulation [6,8] in terms of the entropy $H(X) = \langle -\log_2(x_{n+1}) \rangle$ and entropy rate [18, 20] $H_\mu(X) = \lim_{k \to \infty} \langle -\log_2(x_{n+1} | x_n^{(k)}) \rangle$:

$$A(X) = H(X) - H_\mu(X), \tag{4}$$

$$H_\mu(X) = \lim_{k \to \infty} H(X^{(k+1)}) - H(X^{(k)}), \tag{5}$$

where $H(X^{(k)})$ represents block entropies of the k consecutive states $X^{(k)}$. For the multivariate normal form here, these terms are computed for each node i from the (lagged) autocovariance terms $\Omega(s)_{ii}$. First, we simply have $H(X_i) = \frac{1}{2} \ln (2\pi e |\Omega(0)_{ii}|)$. The block entropies are then computed via autocovariance matrices $M_i(k)$ (which due to stationarity are symmetric Toeplitz matrices):

$$M_i(k) = \begin{pmatrix} \Omega(0)_{ii} & \Omega(1)_{ii} & \Omega(2)_{ii} & \cdots & \Omega(k-1)_{ii} \\ \Omega(1)_{ii} & \Omega(0)_{ii} & \Omega(1)_{ii} & & \Omega(k-2)_{ii} \\ \Omega(2)_{ii} & \Omega(1)_{ii} & \Omega(0)_{ii} & & \vdots \\ \vdots & & & \ddots & \Omega(1)_{ii} \\ \Omega(k-1)_{ii} & & \cdots & & \Omega(0)_{ii} \end{pmatrix},$$

so $H(X_i^{(k)}) = \frac{1}{2} \ln (2\pi e)^k |M_i(k)|$, and [via Eqs. (4) and (5)]

$$A(X_i) = \lim_{k \to \infty} \tfrac{1}{2} \{\ln [|\Omega(0)_{ii}||M_i(k)|/|M_i(k+1)|]\}. \tag{6}$$

Numerically calculating $A(X_i)$ from C is straightforward via Eqs. (2) and (3), $M_i(k)$, and Eq. (6). Use of $k \to \infty$ is correct for this measure in general (unless one can establish finite Markovian dependence of the destination on its past) [6,7], although limited by computational time here.

IV. ANALYTIC CALCULATION

In the spirit of the analysis of Barnett *et al.* [1,2] on the Tononi-Sporns-Edelman (TSE) complexity [3], we now make analytic expansions of Eq. (6) to directly reveal the roles of motif structures in $A(X_i)$. For the purposes of this analysis we assume the nodes do not have self-connections, i.e., $C_{ii} \to 0$ (partially following [1,2]). We treat the full general case in an extended version of this work [21].

We start by writing the relevant autocovariance terms from $M_i(k+1)$. We limit our considerations to be accurate to the highest order of contributions to $A(X_i)$ of motifs of size three [which will be at $O(\epsilon^6)$ in our result in Eq. (12), where $\epsilon \equiv \|C\|$ for any consistent matrix norm [1]]. This expansion favors shorter paths, since stored information is gradually "forgotten"

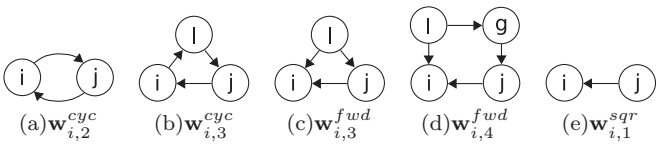

FIG. 1. Motifs implicated in calculation of information storage at node i, including *directed cycles* (a) and (b); *feedforward loop motifs* (c) and (d); and *directed effects* (e).

along the path. Expanding Eq. (2) and Eq. (3) we have

$$\begin{aligned}
\Omega(s)_{ii} &= [(I + C^T C + O(\epsilon^4))C^s]_{ii},\\
\Omega(0)_{ii} &= 1 + \sum_{j\neq i} C_{ji}^2 + O(\epsilon^4) := 1 + \mathbf{w}_{i,1}^{\mathrm{sqr}} + O(\epsilon^4),\\
\Omega(1)_{ii} &= \sum_{j\neq i;l\neq j,i} C_{li}C_{lj}C_{ji} + O(\epsilon^5) := \mathbf{w}_{i,3}^{\mathrm{fwd}} + O(\epsilon^5),\\
\Omega(2)_{ii} &= \sum_{j\neq i} C_{ij}C_{ji} + \sum_{g;l\neq i,g;j\neq i,g} C_{li}C_{lg}C_{gj}C_{ji} + O(\epsilon^6)\\
&:= \mathbf{w}_{i,2}^{\mathrm{cyc}} + \mathbf{w}_{i,4}^{\mathrm{fwd}} + O(\epsilon^6),\\
\Omega(3)_{ii} &= \sum_{j\neq i;l\neq j,i} C_{il}C_{lj}C_{ji} + O(\epsilon^5) := \mathbf{w}_{i,3}^{\mathrm{cyc}} + O(\epsilon^5). \qquad (7)
\end{aligned}$$

$\Omega(s)_{ii}$ for $s \geqslant 4$ will be $O(\epsilon^4)$ or smaller and enter $A(X_i)$ only below $O(\epsilon^6)$ [in our result at Eq. (12)]. We see already the role of loop motifs in establishing the delayed autocovariance terms, with only length s feedback and feedforward loops involved in each $\Omega(s \geqslant 1)_{ii}$ term. Weighted motif sums, e.g., $\mathbf{w}_{i,3}^{\mathrm{fwd}}$, are defined by the summations in the above equations and shown in Fig. 1 (the *network averages* $\mathbf{m}_{2,2}$, $\mathbf{m}_{3,8}$, and $\mathbf{m}_{3,3}$ in [2] are computed over these *local* motif sums $\mathbf{w}_{i,2}^{\mathrm{cyc}}$, $\mathbf{w}_{i,3}^{\mathrm{cyc}}$, and $\mathbf{w}_{i,3}^{\mathrm{fwd}}$).

We turn now to the $\{-\ln[|M_i(k+1)|/|M_i(k)|]\}$ term in $A(X_i)$ in Eq. (6). From [1] we have

$$\ln|R| = \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \mathrm{trace}(\hat{R}^m), \qquad (8)$$

with $\hat{R} = R - I$. We can then write

$$\lim_{k\to\infty} -\ln\frac{|M_i(k+1)|}{|M_i(k)|} = \lim_{k\to\infty} \sum_{m=1}^{\infty} \frac{(-1)^m}{m} r_i(k,m), \qquad (9)$$

$$r_i(k,m) = \mathrm{trace}[\hat{M}_i(k+1)^m] - \mathrm{trace}[\hat{M}_i(k)^m] \qquad (10)$$

and begin to form estimates of $A(X_i)$ by taking Eq. (9) up to particular values of m. Now, since the differences $r_i(k,m)$ must come from the contributions of the extra final row and column of $\hat{M}_i(k+1)^m$, then

$$\begin{aligned}
r_i(k,m) &= \sum_{\substack{q_1,q_2,\ldots,q_m\\ \exists v \mid q_v = k+1}} \hat{M}_i(k+1)_{q_1,q_2}\cdots\hat{M}_i(k+1)_{q_m,q_1},\\
r_i(k,1) &= \hat{M}_i(k+1)_{k+1,k+1} = \Omega(0)_{ii} - 1,\\
r_i(k,2) &= [\Omega(0)_{ii} - 1]^2 + 2\sum_{q=1}^{k} \Omega(q)_{ii}^2,\\
r_i(k,3) &= [\Omega(0)_{ii} - 1]^3 + 6(\Omega(0)_{ii} - 1)\sum_{q=1}^{k} \Omega(q)_{ii}^2 + O(\epsilon^7).
\end{aligned}$$

Using $C_{ii} \to 0$ limits the terms in $r_i(k,3)$ at $O(\epsilon^6)$ and means there are no extra terms at $O(\epsilon^6)$ for $r_i(k,m > 3)$. Now, $r_i(k,m)$ always has a term $[\Omega(0)_{ii} - 1]^m$ (from $q_v = k+1, \forall\, v$), and these cancel with the $\ln|\Omega(0)_{ii}|$ term in Eq. (6). Using the $r_i(k,m)$ expansions above we can form estimates for $A(X_i)$ via Eq. (9), limiting k to capture terms up to $O(\epsilon^4)$ and $O(\epsilon^6)$, respectively:

$$A^*(X_i) = \tfrac{1}{2}\left(\mathbf{w}_{i,2}^{\mathrm{cyc}}\right)^2, \qquad (11)$$

$$\begin{aligned}
A^{**}(X_i) = \tfrac{1}{2}\big[&\mathbf{w}_{i,2}^{\mathrm{cyc}}\left(\mathbf{w}_{i,2}^{\mathrm{cyc}} + 2\mathbf{w}_{i,4}^{\mathrm{fwd}} - 2\mathbf{w}_{i,1}^{\mathrm{sqr}}\mathbf{w}_{i,2}^{\mathrm{cyc}}\right)\\
&+ \left(\mathbf{w}_{i,3}^{\mathrm{fwd}}\right)^2 + \left(\mathbf{w}_{i,3}^{\mathrm{cyc}}\right)^2\big]. \qquad (12)
\end{aligned}$$

Then, we notice that

$$\begin{aligned}
\mathbf{w}_{i,4}^{\mathrm{fwd}} &= \sum_{l\neq i;j\neq i} C_{li}^2 C_{ij} C_{ji} + \sum_{g\neq i;l\neq i,g;j\neq i,g} C_{li}C_{lg}C_{gj}C_{ji},\\
&= \mathbf{w}_{i,1}^{\mathrm{sqr}}\mathbf{w}_{i,2}^{\mathrm{cyc}} + \mathbf{w}_{i,4}^{\mathrm{fwd}'}, \qquad (13)
\end{aligned}$$

with $\mathbf{w}_{i,4}^{\mathrm{fwd}'}$ representing $\mathbf{w}_{i,4}^{\mathrm{fwd}}$ with the additional restriction $g \neq i$. Equation (13) can simplify Eq. (12) as such:

$$A^{**}(X_i) = \tfrac{1}{2}\left[\mathbf{w}_{i,2}^{\mathrm{cyc}}\left(\mathbf{w}_{i,2}^{\mathrm{cyc}} + 2\mathbf{w}_{i,4}^{\mathrm{fwd}'}\right) + \left(\mathbf{w}_{i,3}^{\mathrm{fwd}}\right)^2 + \left(\mathbf{w}_{i,3}^{\mathrm{cyc}}\right)^2\right].$$

$A^*(X_i)$ captures $A(X_i)$ up to the highest-order contribution from two-node motifs, while $A^{**}(X_i)$ is accurate to the highest-order contribution from three-node motifs.

V. FURTHER ANALYSIS AND DISCUSSION

Our estimate $A^{**}(X_i)$ shows that information storage locally at node i is dominated by two types of functional [15] network motif involving that node. The first are *directed cycles* (or feedback loops) $\mathbf{w}_{i,2}^{\mathrm{cyc}}$ and $\mathbf{w}_{i,3}^{\mathrm{cyc}}$, which were hypothesized earlier to enable *distributed* information storage: node i can send information out to its neighbors, then receive it back via these loops at a later time. That is, loops allow information to cycle. The second are *feedforward loop motifs* $\mathbf{w}_{i,3}^{\mathrm{fwd}}$ and $\mathbf{w}_{i,4}^{\mathrm{fwd}'}$: each captures dual paths from node l to i of *different* lengths. This facilitates the arrival of the same information from l at i at two different time steps; i.e., information in i at one time point is effectively being *stored for it* elsewhere in the network before being available again at a later time. These motifs provide a vital mechanism for short-term storage in feedforward networks. Longer motifs of both types will start to appear in higher-order approximations.

Crucially, we have a positive correlation between the information storage capability of each node i and these weighted motif counts for it. Further, if the network has only *positive* edge weights, we have a positive correlation to the *number* of such motifs; and, where the edge weights are positive *on average* (as per the mammalian cortex, where the majority of connections are thought to be excitatory [2]) the same conclusion applies *on average*. These insights have interesting implications: providing analytic evidence why some recurrent neural networks are known for good memory performance [12,13,16], and a computational reason for the relative abundance or retention of reciprocal links ($\mathbf{w}_{i,2}^{\mathrm{cyc}}$) and connected three-node structural motifs [22] in the mammalian cortex [11,17] and of $\mathbf{w}_{i,3}^{\mathrm{fwd}}$ in gene regulatory networks [14].

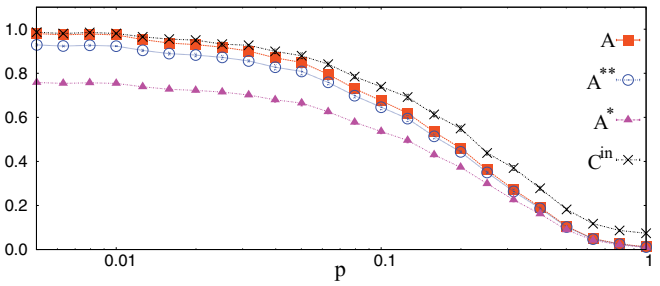

FIG. 2. (Color online) Information storage and relation to clustering coefficient through a small-world transition on an $N = 100$ ring network. Initially each node has $K = 4$ *directed* incoming links from its closest neighbors; then the sources of each link are randomized with probability p. Dynamics are generated using Eq. (1) with equal link weights $c = 0.5/K$. Information storage A (with $k = 30$) and estimates A^* and A^{**} are measured for each node i; then network averages are taken, and all are normalized to $A(p = 0)$. (Vanishing) error bars indicate the standard error of the mean across ten network realizations for each p. Also displayed are network means of clustering coefficients $\tilde{C}_i^{\text{in}}$ for $\mathbf{w}_{i,3}^{\text{fwd}}$ directed motifs, normalized to $\tilde{C}^{\text{in}}(p = 0)$.

To assist further analysis, we present a brief application of $A(X_i)$ and its estimates (computing Ω with [23]) to Watts-Strogatz ring networks [4] in Fig. 2, where the directed links are gradually randomized with probability p and undergo a small-world transition. This is a useful example since increasing p takes us from a regular lattice with many of the motifs of interest to randomized networks where few remain. Figure 2 emphasizes several important points. First, with the positive link weights here, we see a very clear monotonic decrease in $\langle A(X_i)\rangle$ as the motifs are decayed by randomization, as predicted by our estimates. Further, Fig. 2 shows that our estimates provide reasonable approximations to $A(X_i)$ which improve from $A^*(X_i)$ to $A^{**}(X_i)$ as higher-order terms are included. This highlights that the reciprocal links $\mathbf{w}_{i,2}^{\text{cyc}}$ provide the largest storage component here; clearly *undirected networks* will have significantly larger storage capability than directed networks. Further, the contribution of the three-node motifs can be *directly* expressed in terms of weighted *clustering coefficients* for the corresponding motif [24], e.g., we have $\mathbf{w}_{i,3}^{\text{fwd}} = \tilde{C}_i^{\text{in}} K(K-1)c^2$ for this generic case of equal edge weights c. Clearly, nodes with higher clustering coefficients store more information. We explore the relationship between clustered structure and information storage further in [21].

Our estimates are improved of course by accounting for higher-order terms, yet the returns in accuracy for doing so are clearly diminishing. Similarly, our estimates improve as coupling strengths C_{ji} weaken; we explore this accuracy in a forthcoming extension [21].

These analytic results are of course limited to linear interactions only, although we emphasize that they should be seen as "approximations in the weakly coupled near-linear regime of nonlinear dynamics" [1]. Crucially though, the decaying information storage profile with increasing randomness through a small-world transition is very similar to numerical results obtained from random Boolean dynamics [5]. Certainly such Boolean dynamics can hardly be labeled linear, nor as having "positive" weights, and as such provide impetus for the potential generality of our results. As argued in [5], these results suggest an explanation for the dynamic importance of clustered structure in small-world networks.

The decay of correlations within the network is known to be bounded by the dominant eigenvalue λ of C, which plays an important role in the dynamics [19]. Unlike $A(X_i)$ however, λ cannot differentiate the dynamics with p in Fig. 2, since with fixed weighted in-degree cK for all nodes, λ is the same for all networks. More generally, it is easy to produce examples of isospectral networks (i.e., with the same eigenvalues) with differing $\langle A(X_i)\rangle$, showing that $\langle A(X_i)\rangle$ is not directly determined by the eigenvalues. While correlated, there is a conceptual difference between the computational perspective of information storage in the *dynamics of individual nodes* and an eigenvalue inference of *network-wide, persistent* memory: eigenvalues capture persistent memory in feedback loops, but do not capture the transient storage in feedforward motifs (e.g., the network of motif $\mathbf{w}_{i,3}^{\text{fwd}}$ has only zero eigenvalues). We explore this further in [21].

Interestingly, some of the same motifs underpinning information storage here were identified as driving TSE complexity [1,2], although their precise contributions are different. This suggests that TSE complexity contains a significant flavor of information storage capability (aligning with insights from an information-geometric framework [25]), providing additional insight into the decay of TSE complexity through small-world transitions [2].

Our results revealing clustered structure, in particular directed cycles and feedforward loop motifs, as the dominant information storage contributors align with intuition and have interesting implications. In a forthcoming presentation, we extend this result to the more general case involving self-connections [21].

[1] L. Barnett, C. L. Buckley, and S. Bullock, Phys. Rev. E **79**, 051914 (2009).
[2] L. Barnett, C. L. Buckley, and S. Bullock, Phys. Rev. E **83**, 041906 (2011).
[3] G. Tononi, O. Sporns, and G. Edelman, Proc. Natl. Acad. Sci. USA **91**, 5033 (1994).
[4] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[5] J. T. Lizier, S. Pritam, and M. Prokopenko, Artif. Life **17**, 293 (2011).
[6] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, Inf. Sci. **208**, 39 (2012).
[7] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, Phys. Rev. E **77**, 026110 (2008).
[8] J. T. Lizier, M. Prokopenko, and A. Y. Zomaya, Chaos **20**, 037109 (2010).
[9] S. Katare and D. H. West, Complexity **11**, 26 (2006).
[10] M. G. Kitzbichler, M. L. Smith, S. R. Christensen, and E. Bullmore, PLoS Comput. Biol. **5**, e1000314 (2009).

[11] O. Sporns, *Networks of the Brain* (MIT Press, Cambridge, MA, 2011).
[12] J. Boedecker, O. Obst, N. M. Mayer, and M. Asada, HFSP J. **3**, 340 (2009).
[13] H. Jaeger and H. Haas, Science **304**, 78 (2004).
[14] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[15] O. Sporns and R. Kötter, PLoS Biol. **2**, e369 (2004).
[16] J. L. Elman, Cognit. Sci. **14**, 179 (1990).
[17] S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii, PLoS Biol. **3**, e68 (2005).
[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, New York, 1991).
[19] S. S. Soliman and M. D. Srinath, *Continuous and Discrete Signals and Systems* (Prentice-Hall, Englewood Cliffs, NJ, 1998)
[20] J. P. Crutchfield and D. P. Feldman, Chaos **13**, 25 (2003).
[21] J. T. Lizier, F. M. Atay, and J. Jost (unpublished).
[22] Connected three-node *structural* motifs must include one of the *functional* motifs [15] $\mathbf{w}_{i,3}^{\mathrm{cyc}}$ or $\mathbf{w}_{i,3}^{\mathrm{fwd}}$.
[23] L. Barnett, C. L. Buckley, and S. Bullock, Matlab routines for neural complexity analysis (2009), http://www.secse.net/ncomp/ncomp_tools.zip
[24] G. Fagiolo, Phys. Rev. E **76**, 026107 (2007).
[25] N. Ay, E. Olbrich, N. Bertschinger, and J. Jost, Chaos **21**, 037103 (2011).