

Research Article

The Emergence of Informative Higher Scales in Complex Networks

Brennan Klein ^{1,2} and Erik Hoel ³

¹Network Science Institute, Northeastern University, Boston, MA 02115, USA

²Laboratory for the Modeling of Biological and Socio-Technical Systems, Northeastern University, Boston, MA 02115, USA

³Allen Discovery Center, Tufts University, Medford, MA 02155, USA

Correspondence should be addressed to Erik Hoel; hoelerik@gmail.com

Received 11 October 2019; Revised 8 January 2020; Accepted 21 January 2020; Published 4 April 2020

Academic Editor: Yong Xu

Copyright © 2020 Brennan Klein and Erik Hoel. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The connectivity of a network contains information about the relationships between nodes, which can denote interactions, associations, or dependencies. We show that this information can be analyzed by measuring the uncertainty (and certainty) contained in paths along nodes and links in a network. Specifically, we derive from first principles a measure known as *effective information* and describe its behavior in common network models. Networks with higher effective information contain more information in the relationships between nodes. We show how subgraphs of nodes can be grouped into macronodes, reducing the size of a network while increasing its effective information (a phenomenon known as *causal emergence*). We find that informative higher scales are common in simulated and real networks across biological, social, informational, and technological domains. These results show that the emergence of higher scales in networks can be directly assessed and that these higher scales offer a way to create certainty out of uncertainty.

1. Introduction

Networks provide a powerful syntax for representing a wide range of systems, from the trivially simple to the highly complex [1–3]. It is common to characterize networks based on structural properties like their degree distribution or clustering, and the study of such properties has been crucial for the growth of Network Science. Yet there remains a gap in our treatment of the information contained in the relationships between nodes in a network, particularly in networks that have both weighted connections and feedback, which are hallmarks of complex systems [4, 5]. As we will show, analyzing this information allows for modeling the network at the most appropriate, informative scale. This is especially critical for networks that describe interactions or dependencies between nodes such as contact networks in epidemiology [6], neuronal and functional networks in the brain [7], or interaction networks among cells, genes, or drugs [8], as these networks can often be analyzed at multiple different scales.

Here we introduce information-theoretic measures that capture the information contained in the connectivity of a network, which can be used to identify when these networks possess informative higher scales. To do so, we focus on the out-weight vector, W_i^{out} , of each node, v_i , in a network. This vector consists of weights w_{ij} between v_i and its neighbors, v_j , and $w_{ij} = 0$ if there is no edge from v_i to v_j . For each W_i^{out} we assume $\sum_j w_{ij} = 1$, which means w_{ij} can be interpreted as the probability p_{ij} that a random walker on v_i will transition to v_j in the next timestep, where a random walker might represent the passing of a signal, an interaction, or a state-transition [9]. The information contained in a network's connectivity can be characterized by the uncertainty among its nodes' out-weights and in-weights. The total information in the relationships between nodes is a function of this uncertainty and can be derived from two properties.

The first is the uncertainty of a node's outputs, which is the Shannon entropy [10] of its out-weights, $H(W_i^{\text{out}})$. The average of this entropy, $\langle H(W_i^{\text{out}}) \rangle$, across all nodes is the

amount of noise present in the network's relationships, only if $\langle H(W_i^{\text{out}}) \rangle = 0$ is the network is *deterministic*.

The second property is how weight is distributed across the whole network, $\langle W_i^{\text{out}} \rangle$. This vector is composed of elements that are the sum of the in-weights w_{ji} to each node v_i from each of its incoming neighbors, v_j (then normalized by total weight of the network). Its entropy, $H(\langle W_i^{\text{out}} \rangle)$, reflects how certainty is distributed across the network. If all nodes link only to the same node, then $H(\langle W_i^{\text{out}} \rangle) = 0$, and the network is totally *degenerate* since all nodes lead to the same node.

The *effective information* (*EI*) of a network is the difference between these two quantities:

$$EI = H(\langle W_i^{\text{out}} \rangle) - \langle H(W_i^{\text{out}}) \rangle. \quad (1)$$

The entropy of the distribution of out-weights in the network forms an upper bound of the amount of unique information in the network's relationships, from which the information lost due to the uncertainty of those relationships is subtracted. Networks with high *EI* contain more certainty in the relationships between nodes in the network (since the links represent less uncertain dependencies, unique associations, or deterministic transitions), whereas networks with low *EI* contain less certainty. Note that *EI* can be interpreted simply as a structural property of random walkers on a network and their behavior, similar to other common network measures [9].

Here, we use this measure to develop a general classification of networks (key terms can be found in Supplementary Materials, SM V A). Furthermore, we show how the connectivity and different growth rules of a network have a deep relationship to that network's *EI*. This also provides a principled means of quantifying the amount of information among the micro-, meso-, and macroscale dependencies in a network. We introduce a formalism for finding and assessing the most informative scale of a network: the scale that minimizes the uncertainty in the relationships between nodes. For some networks, a macroscale description of the network can be more informative in this manner, demonstrating a phenomenon known as *causal emergence* [11, 12], which here we generalize to complex networks. This provides a rigorous means of identifying when networks possess an informative higher scale.

2. Results

2.1. Effective Information Quantifies a Network's Dependencies. This work expands to networks previous research on using effective information to measure the amount of information in the causal relationships between the mechanisms or states of a system. Originally, *EI* was introduced to capture the causal influence between two subsets of neurons as a step in the calculation of integrated information in the brain [13]. Later, a system-wide version of *EI* was shown to capture fundamental causal properties in Boolean networks of logic gates, particularly their determinism and degeneracy [11].

Our current derivation from first principles of an *EI* for networks is equivalent to this system-wide definition (SM

V B), which was based originally on interventions upon system states. For example, if a system in a particular state, *A*, always transitions to state *B*, the causal relationship between *A* and *B* can be represented by a node-link diagram wherein the two nodes—*A* and *B*—are connected by a directed arrow, indicating that *B* depends on *A*. This might be a node pair in a “causal diagram” (often represented as a directed acyclic graph, or a DAG) such as those used in [14, 15] to represent interventions and causal relationships. In such a case, the information in the causal relationship between *A* and *B* can be assessed by intervening to randomize *A* ($\text{do}(A = H^{\text{max}})$) and measuring the effects on *B*. The *EI* would be the mutual information between *A* and *B* under such randomization: $I(\text{do}(A = H^{\text{max}}), B)$ [16].

To expand this framework to networks in general, we relax this intervention requirement by assuming that the elements in W_i^{out} sum to 1. In this case, an “intervention” can be interpreted as dropping a random walker on the network. For example, if the network represents a DAG or Markov chain, then dropping a random walker on a node v_i would be equivalent to $\text{do}(v_i)$. The entropy of the transitions of the random walkers and how those transitions are distributed defines the *EI* of a network. In this generalized formulation, only in networks where the nodes and edges actually represent dynamics, interactions, or couplings does *EI* indicate information about causation. In the case where edges represent correlations, or where what nodes or edges represent is undefined, *EI* is merely a structural property of the information contained in the behavior of hypothetical random walkers (however, this situation is no different from other analysis methods that rely on random walkers).

Here we describe how this generalized structural *EI* property behaves in common network models, asking basic questions about the relationship between a network's *EI* and its size, density, and structure. These inquiries allow for the exhaustive classification and quantification of the information contained in the connectivity of real networks. It is intuitive that the *EI* of a network will increase as the network grows in size. In general, adding more nodes should increase the entropy, which should in turn increase the amount of information. However, in cases of randomness rather than structure, *EI* should reflect this randomness. We found this is indeed the case.

Figure 1(a) shows the relationship between a network's *EI* and its size under several parameterizations of Erdős-Rényi (ER) random graphs [17, 18]. As the size of an ER network increases (while keeping constant the probability that any two nodes will be connected, p), its *EI* converges to a value of $-\log_2(p)$. That is, in random networks, *EI* is dominated solely by the probability that any two nodes are connected, a key finding which demonstrates that, after a certain point, a random network structure does not contain more information as its size increases. This shift occurs in ER networks at approximately $\langle k \rangle = \log_2(N)$, which is also the point at which we can expect all nodes to be in a giant component [1]. This finding illustrates that network connectivity must be nonrandom to increase the amount of information in the relationships between nodes (see SM V C 1 for derivation). Note that if a network is maximally dense

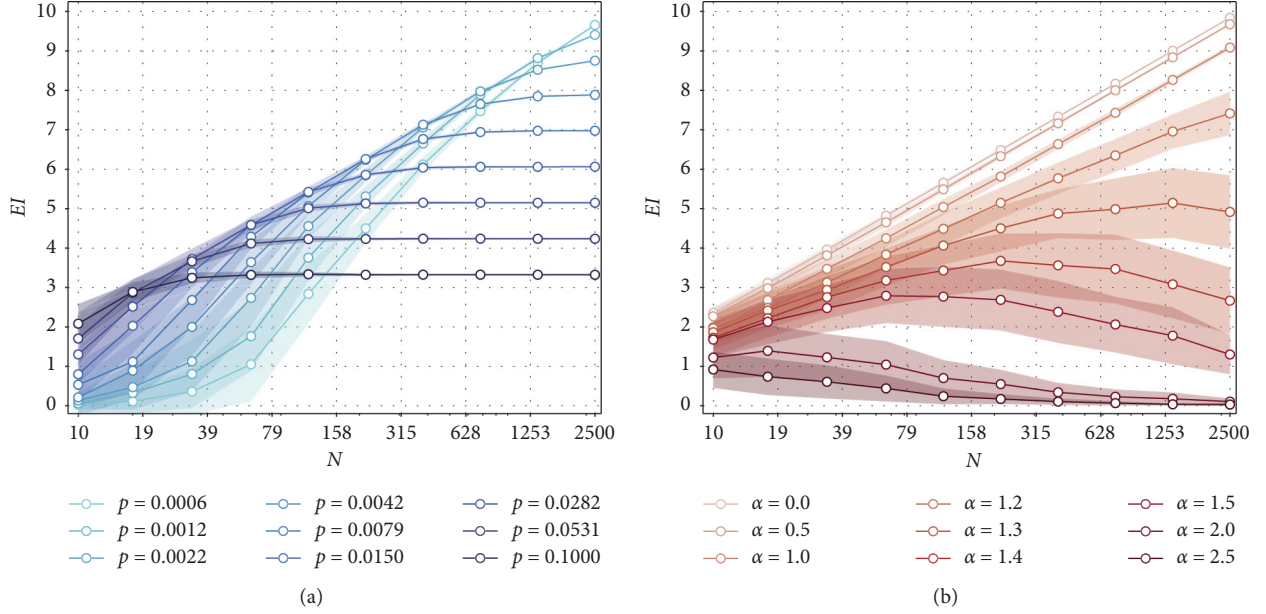


FIGURE 1: Effective information depends on network structure. (a) In Erdős-Rényi (ER) networks, we see the network’s EI level off at $EI = -\log_2(p)$ as N , the network’s size, increases (log scale shown). (b) The EI of networks grown under a preferential attachment mechanism, which depends on the preferential attachment exponent, α . Under this network growth model, new nodes add their m edges (here, $m = 1$) to existing nodes in the network with a probability proportional to k^α . Only sublinear preferential attachment ($\alpha < 1.0$) allows for the continuous growth of EI with the growth of the network. The ribbons around the data represent standard deviations after 100 simulations of each.

(i.e., a fully connected network, with self-loops), $EI = 0.0$. However, we expect such dense low- EI structures to be uncommon, since network structures found in nature and society tend to be sparse [19].

We report another key relationship between a network’s connectivity and its EI in Figure 1(b). We again compare the EI of a network to its size, focusing on networks grown under different parameterizations of a preferential attachment model [20, 21]. Under a preferential attachment growth model, a new node is added to the network at each time step, contributing m new edges to the network; these m edges connect to nodes already in the network, v_j , with a probability proportional to k_j^α . Here, k_j is the degree of node v_j and α tunes the amount of preferential attachment. A value of $\alpha = 0.0$ corresponds to each node having an equal chance of receiving a new node’s link (i.e., no preferential attachment). The classic Barabási-Albert network corresponds to linear preferential attachment, $\alpha = 1.0$ [21]. Superlinear preferential attachment, $\alpha > 1.0$, creates networks that have less and less EI , eventually resembling star-like structures (see SM V C 2 for derivation). As shown in Figure 1(b), only in cases of sublinear preferential attachment, $\alpha < 1.0$, does the network’s EI continue to increase with its size. When $\alpha = 0.0$ —creating a random tree—the network’s EI increases logarithmically as its size increases.

The maximum possible EI in a network of N nodes is $\log_2(N)$. This can be seen in the case of a directed ring network where each node has one incoming link and one outgoing link, each with a weight of 1.0, so each node has one node uniquely connecting to it. In such a network, each node contributes zero uncertainty, since $\langle H(W_i^{\text{out}}) \rangle = 0.0$, and

$H(\langle W_i^{\text{out}} \rangle) = \log_2(N)$, and therefore, its EI is always $\log_2(N)$. In general, the EI of undirected lattices is fixed entirely by its size and the dimension of the ring lattice (i.e., $d = 1$ is an undirected ring, $d = 2$ is a taurus, etc. [22]), so for such lattices $EI = \log_2(N) - \log_2(2d)$ (see SM V C 2 for derivation).

The picture that emerges is that EI is inextricably linked with a network’s connectivity and growth (even network motifs, as shown in SM V D) and therefore to the fundamentals of Network Science. Random networks have a fixed amount of EI , and scale-freeness ($\alpha = 1.0$) represents the critical bound for the growth of EI . In general, dense networks and star-like networks have less EI . The next section explores how EI ’s components explain these associations.

2.2. Determinism and Degeneracy. *Determinism* and *degeneracy* are the two fundamental components of EI [11]. They are based on a network’s connectivity (see Figure 2(a) for a visual explanation), specifically the degree of overlapping weight in the networks. Determinism and degeneracy are derived from the uncertainty over outputs and uncertainty in how those outputs are distributed, respectively:

$$\begin{aligned} \text{determinism} &= \log_2(N) - \langle H(W_i^{\text{out}}) \rangle, \\ \text{degeneracy} &= \log_2(N) - H(\langle W_i^{\text{out}} \rangle). \end{aligned} \quad (2)$$

In a maximally deterministic network wherein all nodes have a single output, $w_{ij} = 1.0$, the determinism is $\log_2(N)$ because $\langle H(W_i^{\text{out}}) \rangle = 0.0$. Conceptually, this means that a

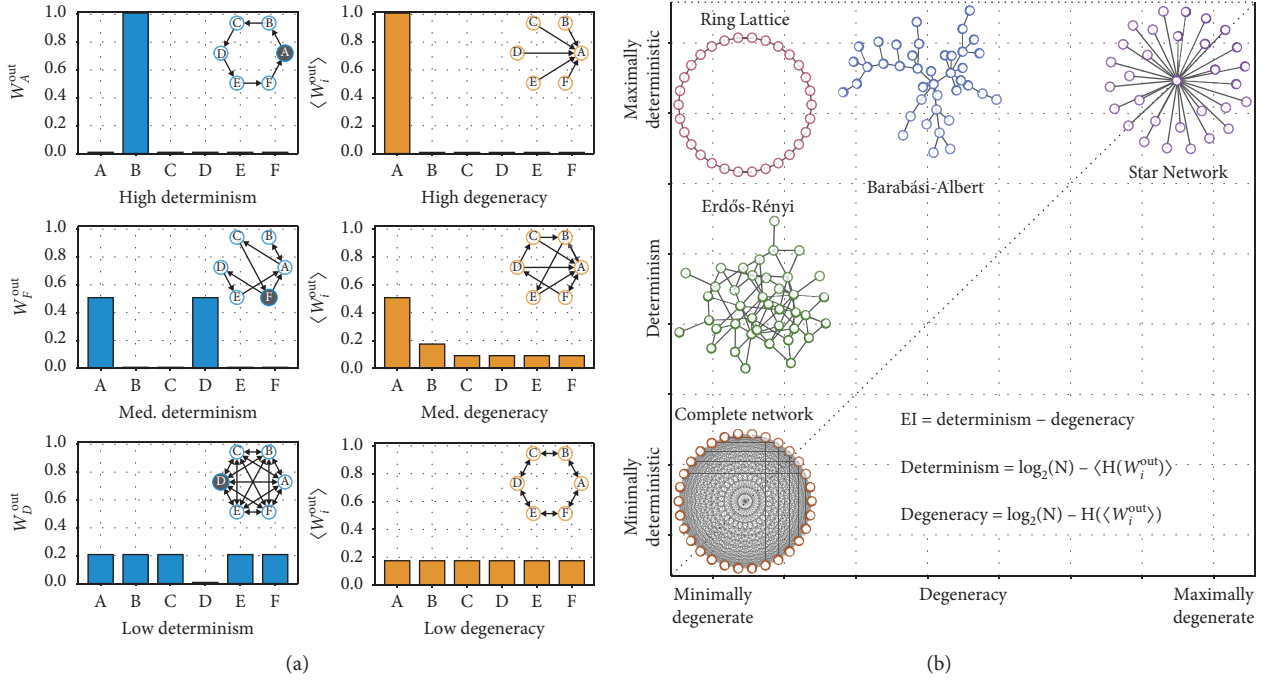


FIGURE 2: Comparing determinism and degeneracy. (a) Left column: three example out-weight vectors, W_i^{out} , of a given node, v_i . A maximally deterministic vector (top left, where the W_A^{out} corresponds to node A in the inset network motif) is when a random walker on v_i transitions to one of its neighbors with probability 1.0, whereas indeterminism occurs when v_i has a uniform probability of visiting any node in the network in the next time step. Right: three example in-weight vectors to a given v_j . A maximally degenerate vector, $\langle W_j^{\text{out}} \rangle$ (top right, exemplified by the inset network motif), is when every outgoing edge in the network connects to a single node, whereas minimal degeneracy occurs when each value in $\langle W_j^{\text{out}} \rangle$ is uniformly $(1/N)$. (b) By comparing the determinism and degeneracy of canonical network structures, we find a great deal of heterogeneity in different network models' ratios between their determinism and degeneracy. High degeneracy is characterized by hub-and-spoke topology, as in the case of the star network. Networks with high determinism are characterized by longer average path lengths, as in the case of a ring lattice.

random walker will move deterministically starting from any node. Degeneracy is the amount of information in the connectivity lost via an overlap in input weights (e.g., if multiple nodes output to the same node). In a perfectly nondegenerate system where all nodes have equal input weights, the degeneracy is zero since $H(\langle W_i^{\text{out}} \rangle) = \log_2(N)$. Together, determinism and degeneracy can be used to define *EI*:

$$EI = \text{determinism} - \text{degeneracy} \quad (3)$$

These two quantities provide clear explanations for why different networks have the *EI* they do. For example, as the size of an Erdős-Rényi random network increases, its degeneracy approaches zero, which means the *EI* of a random network is driven only by the determinism of the network, which is in turn the negative log of the probability of connection, p . Similarly, in d -dimensional ring lattice networks, the degeneracy term is always zero, which means the *EI* of a ring lattice structure also reduces to the determinism of that structure. Ring networks with an average degree $\langle k \rangle$ will have a higher *EI* than ER networks with the same average degree because ring networks will have a higher determinism value. In the case of star networks, the degeneracy term alone governs the decay of the *EI* such that hub-and-spoke-like structures quickly become

uninformative in terms of cause and effect (see SM V C for derivations concerning these cases). In general, this means that canonical networks can be characterized by their ratio of determinism to degeneracy (see Figure 2(b)).

2.3. Effective Information in Real Networks. So far, we have been agnostic as to the origin of the network under analysis. As described previously, to measure the *EI* of a network, one can create each W_i^{out} by normalizing each node's out-weight vector to sum to 1.0. Regardless of what the relationships between the nodes represent, the network's determinism reflects how targeted the out-weights of the nodes are (networks with more targeted links possess higher *EI*), while the degeneracy captures the overlap of the targeting of nodes. High *EI* reflects the greater specificity in the connectivity, whereas low *EI* indicates a lack of specificity (as in Figure 2(a)). This generalizes our results to multiple types of representations, although the origin of the normalized network should be kept in mind when interpreting the value of the measure.

Since the *EI* of a network will change depending on the network's size, we use a normalized form of *EI* known as *effectiveness* in order to compare the *EI* of real networks. Effectiveness ranges from 0.0 to 1.0 and is defined as

$$\text{effectiveness} = \frac{EI}{\log_2(N)}. \quad (4)$$

As the determinism and degeneracy of a network increase to their minimum and maximum possible values, respectively, the effectiveness of that network will trend to 0.0. Regardless of its size, a network wherein each node has a deterministic output to a unique target has an effectiveness of 1.0.

In Figure 3, we examine the effectiveness of 84 different networks corresponding to data from real systems. These networks were selected primarily from the Konect Network Database [23], which was used because its networks are publicly available, range in size from dozens to tens of thousands of nodes, often have a reasonable interpretation as being based on interactions between nodes, and they are diverse, ranging from social networks, to power networks, to metabolic networks. We defined four categories of interest: biological, social, informational, and technological. We selected our networks by using all the available networks (under 40,000 nodes due to computational constraints) in the domains corresponding to each category within the Konect database, and where it was appropriate, the Network Repository as well [24]. See Materials & Methods section and SM Table II for a full description of this selection process.

Lower effectiveness values correspond to structures that have either high degeneracy (as in right column, Figure 2(a)) or low determinism (as in left column, Figure 2(a)) or a combination of both. In the networks we measured, biological networks on average have lower effectiveness values, whereas technological networks on average have the highest effectiveness. This finding aligns intuitively with what we know about the relationship between EI and network structure, and it also supports long-standing hypotheses about the role of redundancy, degeneracy, and noise in biological systems [25, 26]. On the other hand, technological networks like power grids, autonomous systems, or airline networks on average are associated with higher effectiveness values. One explanation for this difference is that efficiency in human-made technological networks tends to create sparser, nondegenerate networks with higher effectiveness on average, wherein the nodes relationships are more specific in their targeting.

Perhaps it might be surprising to find that evolved networks have such low effectiveness. But, as we will show, a low effectiveness can actually indicate that there is informative higher-scale (macroscale) connectivity in the system. That is, a low effectiveness can reflect the fact that biological systems often contain higher-scale structure, which we demonstrate in the following section.

2.4. Causal Emergence in Complex Networks. This new global network measure, EI , offers a principled way to answer an important question: what is the scale that best captures the connectivity of a complex system? The resolution to this question is important because science analyzes the structure of different systems at different spatiotemporal scales, often preferring to intervene and observe systems at levels far

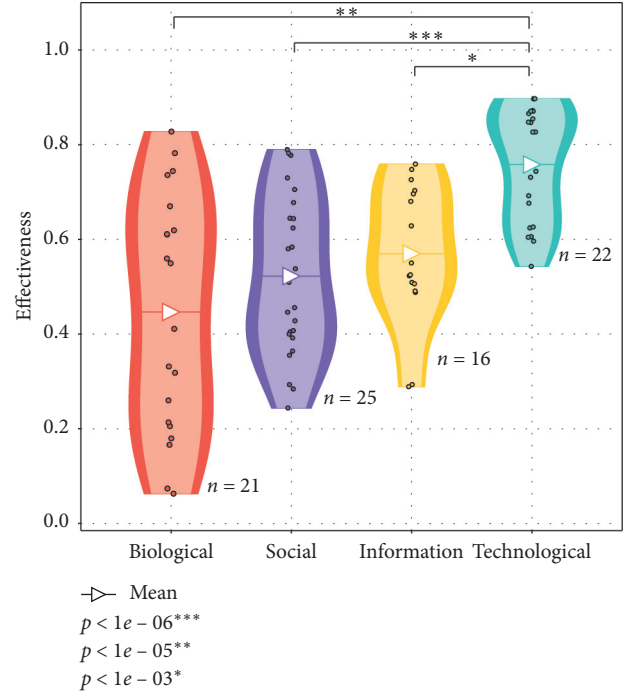


FIGURE 3: Effective information of real networks. Effectiveness, a network's EI , normalized by $\log_2(N)$ [11], of 84 real networks from the Konect network database [23], grouped by domain of origin. To look further at the names and domains of the networks in question, see SM V E Networks in different categories have varying effectiveness (t -test, comparison of means).

above that of the microscale [12]. This is likely because relationships at the microscale can be extremely noisy and therefore uninformative, and coarse-graining can minimize this noise [11]. Indeed, this noise minimization is actually grounded in Claude Shannon's noisy-channel coding theorem [10], wherein dimension reductions can operate like codes that use more of a channel's capacity [16]. Higher-level causal relationships often perform error-correction on the lower-level relationships, thus generating extra effective information at those higher scales. Measuring this difference provides a principled means of deciding when higher scales are more informative (emergence) or when higher scales are extraneous, epiphenomenal, or lossy (reduction).

Bringing these issues to network science, we can now ask, what representation will minimize the uncertainty present in a network? We do this by examining *causal emergence*, which is when a dimensionally reduced network contains more informative connectivity, in the form of a higher EI than the original network. Note that, as discussed, EI can be interpreted solely as a general structural property of networks. Therefore, while we still call this phenomenon "causal emergence" because it has the same mathematical formalization as previous work in Boolean networks and Markov chains [11, 12, 16], here we focus on how it can be used to identify the informative higher scales of networks regardless of what those networks represent.

Notably, the phenomenon can be measured by recasting networks at higher scales and observing how the EI changes,

a process which identifies whether the network’s higher scales add information above and beyond lower scales.

2.5. Network Macroscales. First, we must introduce how to recast a network, G , at a higher scale. This is represented by a new network, G_M . Within G_M , a micronode is a node that was present in the original G , whereas a macronode is defined as a node, μ , that represents a subgraph, S_i , from the original G (replacing the subgraph within the network). Since the original network has been dimensionally reduced by grouping nodes together, G_M will always have fewer nodes than G .

A macronode μ is defined by some W_μ^{out} , derived from the edge weights of the various nodes within the subgraph it represents. One can think of a macronode as being a summary statistic of the underlying subgraph’s behavior, a statistic that takes the form of a single node. Ultimately there are many ways of representing a subgraph, that is, building a macronode, and some ways are more consistent than others in capturing the subgraph’s behavior, depending on the connectivity. We highlight here that macroscales of networks should in general be *consistent* with their underlying microscales in terms of their dynamics. While this has never been assessed within networks or systems generally, there has been previous research that has asked whether the macroscales of structural equation models are consistent with the effect of all possible interventions [27].

Here, to decide whether or not a macronode is an consistent summary of its underlying subgraph, we formalize consistency as a measure of whether random walkers behave identically on G and G_M . We do this because random walks are often used to represent dynamics on networks [9], and therefore, many important analyses and algorithms—such as PageRank for determining a node’s centrality [28] or InfoMap for community discovery [29]—are based on random walks.

Specifically, we define the *inconsistency* of a macroscale as the Kullback-Leibler divergence [30] between the expected distribution of random walkers on G vs. G_M , given some identical initial distribution on each. The expected distribution over G at some future time, t , is $P_m(t)$, while the distribution over G_M at some future time t is $P_M(t)$. To compare the two, the distribution $P_m(t)$ is summed over the same nodes in the macroscale G_M , resulting in the distribution $P_{M|m}(t)$ (the microscale given the macroscale). We can then define the macroscale inconsistency over some series of timesteps T as

$$\text{inconsistency} = \sum_{t=0}^T D_{KL}[P_m(t) \| P_{M|m}(t)]. \quad (5)$$

This consistency measure addresses the extent to which a random dynamical process on the microscale topology will be recapitulated on a dimensionally reduced topology (for how this is applied in our analysis, see Materials & Methods).

What constitutes a consistent macroscale depends on the connectivity of the subgraph that gets grouped into a macronode, as shown in Figure 4. The W_μ^{out} can be

constructed based on the collective W^{out} of the subgraph (shown in Figure 4(a)). For instance, in some cases, one could just coarse-grain a subgraph by using its average W^{out} as the W_μ^{out} of some new macronode μ (as in Figure 4(b)). However, it may be that the subgraph has dependencies not captured by such a coarse-grain. Indeed, this is similar to the recent discovery that when constructing networks from data, it is often necessary to explicitly model higher-order dependencies by using higher-order nodes so that the dynamics of random walks to stay true to the original data [31]. We therefore introduce *higher-order macronodes* (HOMs), which draw on similar techniques to consistently represent subgraphs as single nodes [31].

Different subgraph connectivities require different types of HOMs to consistently represent them. For instance, HOMs can be based on the input weights to the macronode, which take the form $\mu | j$. In these cases, $W_{\mu|j}^{\text{out}}$ is a weighted average of each node’s W^{out} in the subgraph, where the weight is based on the input weight to each node in the subgraph (Figure 4(c)). Another type of HOM that generally leads to consistent macronodes over time is when W_μ^{out} is based on the stationary output from the subgraph to the rest of the network, which we represent as $\mu | \pi$ (Figure 4(d)). These types of HOMs may have minor inconsistencies given some initial state, but will almost always trend toward perfect consistency as the network approaches its stationary dynamics (outlined in Section 4).

Subgraphs with complex internal dynamics can require a more complex type of HOM in order to preserve the macronode’s consistency. For instance, in cases where subgraphs have a delay between their inputs and outputs, this can be represented by a combination of $\mu | j$ and $\mu | \pi$, which when combined captures that delay (Figure 4(e)). In these cases, the macronode μ has two components, one of which acts as a buffer over a timestep. This means that macronodes can possess memory even when constructed from networks that are at the microscale memoryless, and in fact, this type of HOM is sometimes necessary to consistently capture the microscale dynamics.

We present these types of macronodes not as an exhaustive list of all possible HOMs, but rather as examples of how to construct higher scales in a network by representing subgraphs as nodes and also sometimes using higher-order dependencies to ensure those nodes are consistent. This approach offers a complete generalization of previous work on coarse-grains [11] and also black boxes [16, 32, 33], while simultaneously solving the previously unresolved issue of macroscale consistency by using higher-order dependencies. The types of macronodes formed by subgraphs also provide substantive information about the network, such as whether the macroscale of a network possesses memory or path-dependency.

2.6. Causal Emergence Reveals the Scale of Networks. A network has an informative macroscale when a recast network, G_M (a macroscale), has more *EI* than the original network, G (the microscale). In general, networks with lower effectiveness (low *EI* given their size) have a higher potential

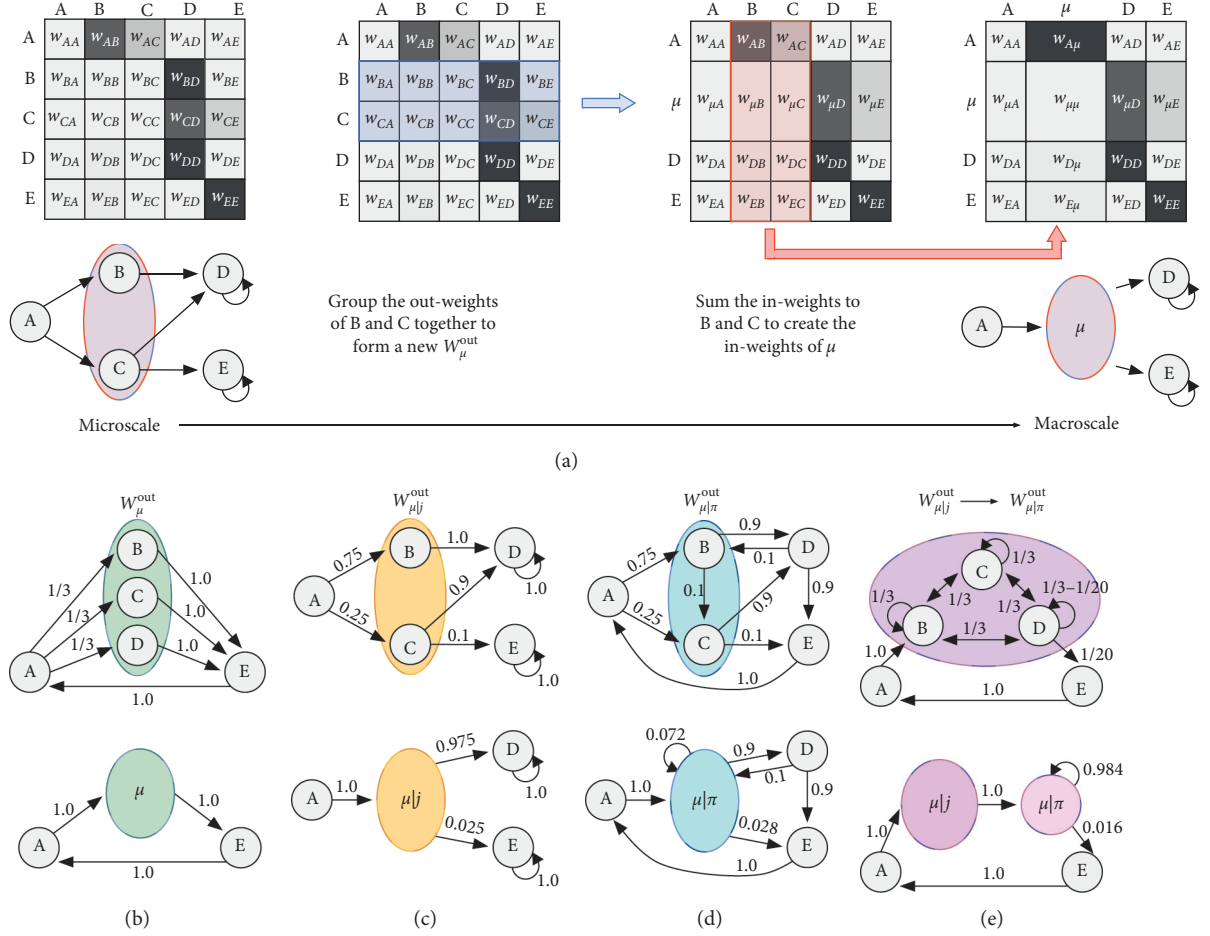


FIGURE 4: Macronodes. (a) The original network, G along with its adjacency matrix (left). The shaded oval indicates that subgraph S member nodes v_B and v_C will be grouped together, forming a macronode, μ . All macronodes are some transformation of the original adjacency matrix via recasting it as a new adjacency matrix (right). The manner of this recasting depends on the type of macronode. (b) The simplest form of a macronode is when W_{μ}^{out} is an average of the W_i^{out} of each node in the subgraph. (c) A macronode that represents some path-dependency, such as input from A. Here, in averaging to create the W_{μ}^{out} , the out-weights of nodes v_B and v_C are weighted by their input from v_A . (d) A macronode that represents the subgraph's output over the network's stationary dynamics. Each node has some associated π_i , which is the probability of v_i in the stationary distribution of the network. The W_{μ}^{out} of a $\mu|\pi$ macronode is created by weighting each W_i^{out} of the micronodes in the subgraph S by $(\pi_i/\sum_{k \in S} \pi_k)$. (e) A macronode with a single timestep delay between input $\mu|j$ and its output $\mu|\pi$, each constructed using the same techniques as its components. However, $\mu|j$ always deterministically outputs to $\mu|\pi$. See SM V A for the full equations governing the creation of the W_{μ}^{out} of each of the different HOMs shown.

for such emergence, since they can be recast to reduce their uncertainty. Searching across groupings allows the identification or approximation of a macroscale that maximizes the EI .

Checking all possible groupings is computationally intractable for all but the smallest networks. Therefore, in order to find macronodes which increase the EI , we use a greedy algorithm that groups nodes together and checks if the EI increases. By choosing a node and then pairing it iteratively with its surrounding nodes we can grow macronodes until pairings no longer increase the EI , and then move on to a new node (see the Materials & Methods section for details on this algorithm).

By generating undirected preferential attachment networks and varying the degree of preferential attachment, α , we observe a crucial relationship between preferential

attachment and causal emergence. One of the central results in network science has been the identification of “scale-free” networks [21]. Our results show that networks that are not “scale-free” can be further separated into micro-, meso-, and macroscales depending on their connectivity. This scale can be identified based on their degree of causal emergence (Figure 5(a)). In cases of sublinear preferential attachment ($\alpha < 1.0$), networks lack higher scales. Linear preferential attachment ($\alpha = 1.0$) produces networks that are scale-free, which is the zone of preferential attachment right before the network develops higher scales. Such higher scales only exist in cases of superlinear preferential attachment ($\alpha > 1.0$). And past $\alpha > 3.0$ the network begins to converge to a macroscale where almost all the nodes are grouped into a single macronode. The greatest amount of causal emergence is found in mesoscale networks, which is when α is between

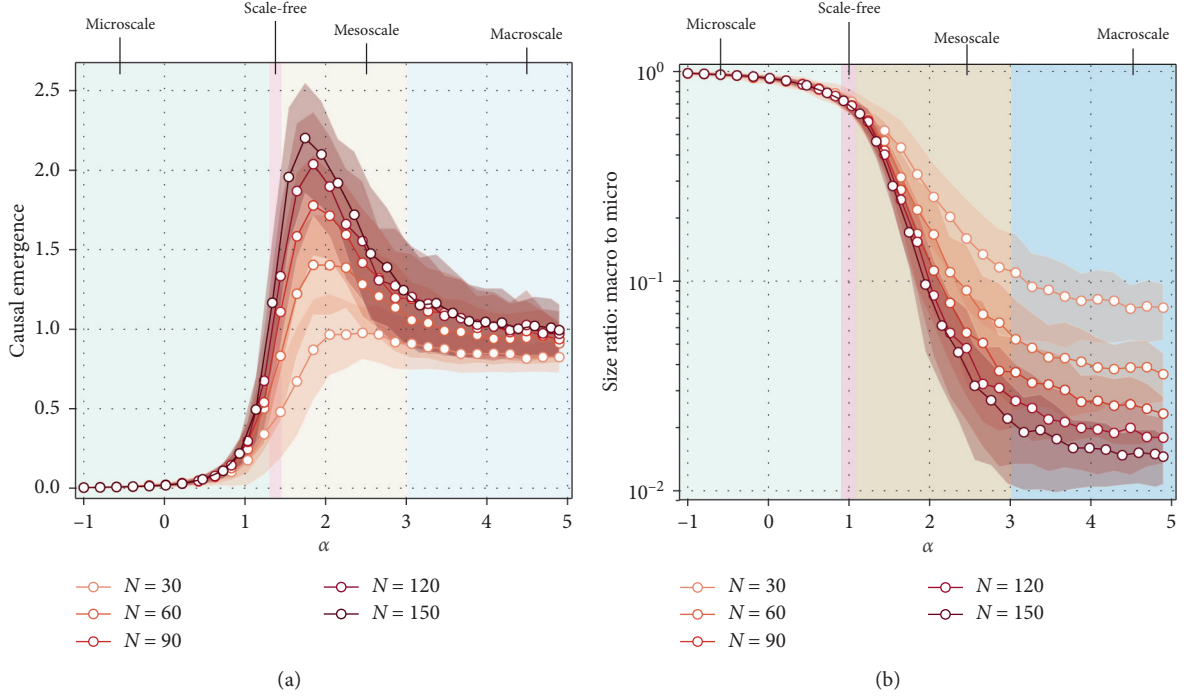


FIGURE 5: The emergence of scale in preferential attachment networks. (a) By repeatedly simulating networks with different degrees of preferential attachment (α values) with $m = 1$ new edge per each new node, and running them through a greedy algorithm (described in Materials & Methods), we observe a distinctive peak of causal emergence once the degree of preferential attachment is above $\alpha = 1$, yielding networks that are no longer “scale-free.” (b) The log of the ratio of original network size, N , to the size of the macroscale network, N_M . Networks with higher α values—more star-like networks—show drastic dimension reductions, and in fact, all eventually reach the same N_M of 2. Comparatively, random trees ($\alpha = 0.0$) show essentially no informative dimension reductions.

1.5 and 3.0, when networks possess a rich array of macronodes. Note that the increase in EI following macroscale groupings for $\alpha > 1.0$ shown in Figure 5(a) resembles the decrease in EI with higher α that we observe in Figure 1(b). This is because after $\alpha > 1.0$ the decreasing EI of the microscale leaves room for improvement of the EI at the macroscale, following a grouping of nodes.

Correspondingly the size of G_M decreases as α increases and the network develops an informative higher scale, which can be seen in the ratio of macroscale network size, N_M , to the original network size, N (Figure 5(b)). As discussed previously, networks generated with higher values for α will be more and more star-like. Star-like networks have higher degeneracy and thus less EI , and because of this, we expect that there are more opportunities to increase the network’s EI through grouping nodes into macronodes. Indeed, the ideal grouping of a star network is when $N_M = 2$ and $EI = 1$ bit. This result is similar to recent advances in spectral coarse-graining that also observe that the ideal coarse-graining of a star network is to collapse it into a two-node network, grouping all the spokes into a single macronode [34], which is what happens to star networks that are recast as macroscales.

Our results offer a principled and general approach to such community detection by asking whether there is an informational gain from replacing a subgraph with a single node. Therefore, we can define *causal communities* as being when a cluster of nodes, or some subgraph, forms a viable

macronode (note that this assumes the connections in the network actually represent possible causal interactions, but it also merely a topological property). Fundamentally, causal communities represent noise at the microscale. The closer a subgraph is to complete noise, the greater the gain in EI by replacing it with a macronode (see SM V G). Minimizing the noise in a given network also identifies the optimal scale to represent that network. However, there must be some structure that can be revealed by noise minimization in the first place. In cases of random networks that form a single large component which lacks any such structure, causal emergence does not occur (as shown in SM V G).

2.7. Causal Emergence in Real Networks. The presence and informativeness of macroscales should vary across real networks, depending on connectivity. Here, we investigate the disposition toward causal emergence of real networks across different domains. We draw from the same set of networks that are analyzed in Figure 3, the selection process and details of which is outlined in the Materials & Methods section. The network sizes span up to 40,000 nodes, thus making it unfeasible to find the best macroscales for each of them. Therefore, we focus specifically on the two categories that previously showed the greatest divergence in terms of the EI : biological and technological. Since we are interested in the general question of whether biological or technological networks show a greater disposition or propensity for

causal emergence, we approximate causal emergence by calculating the causal emergence of sampled subgraphs of growing sizes. Each sample is found using a “snowball sampling” procedure, wherein a node is chosen randomly and then a weakly connected subgraph of a specified size is found around it [35]. This subgraph is then analyzed using the previously described greedy algorithmic approach to find macronodes that maximized the EI in each network. Each available network is sampled 20 times for each size taken from it. In Figure 6, we show how the causal emergence of these real networks differentiates as we increase the sampled subgraph size, in a sequence of 50, 100, 150, and finally 200 nodes per sample. Networks of these sizes previously provided ample evidence of causal emergence in simulated networks, as in Figure 5(a). Comparing the two categories of real networks, we observe a significantly greater propensity for causal emergence in biological networks, and that this is more articulated the larger the samples are. Note that constructing a random null model of these networks (e.g., a configuration model) would tend to create networks with minimal or negligible causal emergence, as is the case for ER networks (Figure 13 in SM V G).

That subsets of biological systems show a high disposition toward causal emergence is consistent, and even explanatory, of many long-standing hypotheses surrounding the existence of noise and degeneracy in biological systems [36]. It also explains the difficulty of understanding how the causal structure of biological systems functions, since they are cryptic by containing certainty at one level and uncertainty at another.

3. Discussion

We have shown that the information in the relationships between nodes in a network is a function of the uncertainty intrinsic to their connectivity as well as how that uncertainty is distributed. To capture this information, we adapted a measure, effective information (EI), for use in networks and analyzed what it reveals about common network structures that have been studied by network scientists for decades. For example, the EI of an ER random network tends to $-\log_2(p)$, and whether the EI of a preferential attachment network grows or shrinks as new nodes are added is a function of whether its degree of preferential attachment, α , is greater or less than 1.0. In networks where the mechanisms or transitions are unknown, but the structure is known, EI captures the degree of unique targeting in the network. In real networks, we showed that the EI of biological networks tends to be much lower than technological networks.

We also illustrated that what has been called “causal emergence” can occur in networks. This is the gain in EI that occurs when a network, G , is recast as a new network, G_M . Finding this sort of informative higher scale means balancing the minimization of uncertainty while simultaneously maximizing the number of nodes in the network. These methods may be useful in improving scientific experimental design, the compression and search of big data, model choice, and even machine learning. Importantly, not

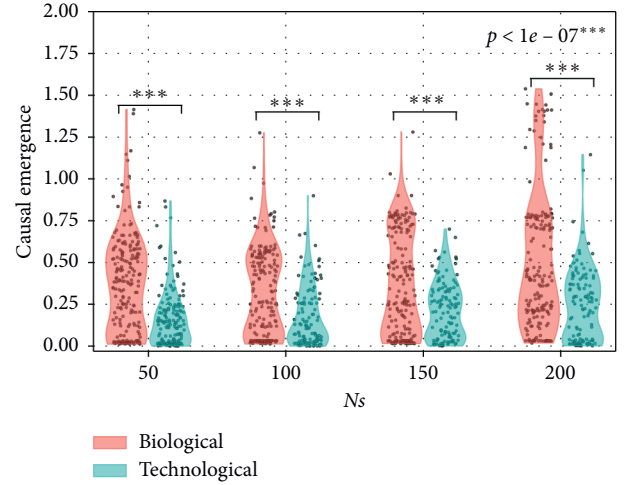


FIGURE 6: Propensity for causal emergence in real networks. Growing snowball samples of the two network domains that previously showed the greatest divergence in *effectiveness*: technological and biological networks. At each snowball size, N_s , each network is sampled 20 times. Across these samples, the total amount of causal emergence for a given sample size is significantly different between the two domains (t -test, comparison of means).

every recast network, G_M , will have a higher EI than the G that it represents, that is, these same techniques can identify cases of reduction. Ultimately, this is because comparing the EI of different network representations provides a ground for comparing the effectiveness of any two network representations of the same complex system. These techniques allow for the formal identification of the scale of a network. Scale-free networks can be thought of as possessing a fractal pattern of connectivity [37], and our results show that the scale of a network is the breaking of that fractal in one direction or the other. Note that a future area of research is how to efficiently identify such informative higher scales, as well as how network properties beyond the EI change across scales [38].

The study of higher-order structures in networks is an increasingly rich area of research [29, 39–42], often focusing on constructing networks that better capture the data they represent. Here, we introduce a formal and generalized way to recast networks at a higher scale while preserving random walk dynamics. In many cases, a macroscale of a network can be just as consistent in terms of random walk dynamics and also possess greater EI . Some macronodes in a macroscale may be of different types with different higher-order properties. In other words, we show how to turn a lower-order network into a higher-order network. One noteworthy and related aspect of our work is demonstrating how a system that is memoryless at the microscale can actually possess memory at the macroscale, indicating that whether a system has memory is a function of scale.

While some [43] have previously recast subgraphs as individual nodes as we do here, they have not done so in ways that are based on noise minimization and maximizing consistency, focusing instead on gains to algorithmic speed via compression. Explicitly creating macronodes to

minimize noise brings the dependencies of the network into focus. This means that causal emergence in networks has a direct relationship to community detection, a vast subdiscipline that treats dense subgraphs within a network as representing shared properties, membership, or functions [44, 45]. However, the relationship between causal emergence and traditional community detection is not as direct as it may seem. For one, causal emergence is high in networks with high degeneracy (i.e., networks with high-degree hubs, as we show in Figure 5(a)). Community detection algorithms do not typically select for such structural properties, instead focusing on dense subgraphs that connect more highly within the subgraph than outside [44]. In SM Figure 12, we show a landscape of stochastic block model networks and their associated values for causal emergence. Indeed in networks that would have high modularity [46] (e.g., two disconnected cliques), we do observe causal emergence, but only when the two disconnected cliques are of *different sizes*. This distinction is key and situates networks that display causal emergence in a meaningful place in the study of complex networks. In light of this, macronodes offer a sort of community detection where the micronodes that make up a macronode are a community and ultimately can be replaced by a macronode that summarizes their behavior while reducing the subgraph's noise. Under this interpretation, community structure is characterized by noise rather than shared memberships.

4. Materials and Methods

4.1. Selection of Real Networks. Networks were chosen to represent the four categories of interest: social, informational, biological, and technological (see SM Figure 10, where we detail the same information as in Figure 3, but also include the source of the network data in addition to the effectiveness value of each network). We used all the available networks under 40,000 nodes (due to computational constraints) within all the domains in the Konect database that reflected our categories of interest. For our social category, we used the domains *Human Contact*, *Human Social*, *Social*, and *Communication*. For our information category, we used the domains *Citations*, *Co-authorship*, *Hyperlinks*, *Lexical*, and *Software*. For our biological category we used the domains *Trophic* and *Metabolic*. Due to overlaps between the Konect database and the Network Repository [24] in these domains, and the paucity of other biological data in the Konect database, we also included the *Brains* domain and the *Ecology* domain from the Network Repository to increase our sample size (again, all networks within these domains under 40,000 nodes were included). For our technological category, we used the domains *Computer* and *Infrastructure* from the Konect database. Again due to overlap between the Konect database and the Network Repository, we also included the *Technological* and *Power Networks* domains from the Network Repository. For a full table of the networks used in this study, along with their source and categorization, see Table II.

4.2. Creating Consistent Macronodes. Previously we outlined methods for creating consistent macronodes of different types. Here, we explore their implementation, which requires deciding which macroscales are consistent. Inconsistency is measured as the Kullback-Leibler divergence between the expected distribution of random walkers on both the microscale (G) and the macroscale (G_M), given an initial distribution, as in equation (5).

To measure the inconsistency we use an initial maximum entropy distribution on the shared nodes between G and G_M , that is, only the set of nodes that are left ungrouped in G_M . Similarly, we only analyze the expected distribution over that same set of micronodes. Since such distributions are only over a portion of the network, to normalize each distribution to 1.0, we include a single probability that represents all the nonshared nodes between G and G_M (representing when a random walker is on a macronode).

We focus on the shared nodes between G and G_M for the inconsistency measure because (a) it is easy to calculate which is necessary during an algorithmic search, (b) except for unusual circumstances, the inconsistency over the shared nodes still reflects the network as a whole, and (c) even in cases of the most extreme macroscales (such as when $\alpha > 4$ in Figure 5), there are still nodes shared between G and G_M .

Here, we examine our methods of using higher-order dependencies in order to demonstrate that this creates consistent macronodes. We use 1000 simulated preferential attachment networks, which were chosen as a uniform random sample between parameters $\alpha = 1.0$ and 2.0 , $n = 25$ to 35 , and with either $m = 1$ or 2 . These networks were then grouped via the algorithm described in the following section. All macronodes were of the $\mu|\pi$ type, and their inconsistency was checked over 1000 timesteps. These macronodes generally have consistent dynamics, either because they start that way or because they trend to that over time, and of the 1000 networks, only 4 had any divergence greater than 0 after 1000 timesteps. In Figure 11 in SM V F, we show 15 of these simulated networks, along with their parameters, number of macronodes, and consistencies. Note that even in the cases with early nonzero inconsistency, this is always very low in absolute terms of bits, and of the randomly chosen 15, none do not trend toward consistency over time. In our observations, most macronodes converge before 500 timesteps, so in analyzing the real-world networks using the $\mu|\pi$ macronode, we check all macronodes for consistency and only reject those that are inconsistent at 500 timesteps. More details about the algorithmic approach to finding causal emergence can be found in the following section.

4.3. Greedy Algorithm for Causal Emergence. The greedy algorithm used for finding causal emergence in networks is structured as follows: for each node, v_i , in the shuffled node list of the original network, collect a list of neighboring nodes, $\{v_j\} \in B_i$, where B_i is the *Markov blanket* of v_i (in graphical models, the Markov blanket, B_i , of a node, v_i , corresponds to the “parents,” the “children,” and the “parents of the children” of v_i [47]). This means that $\{v_j\} \in B_i$ consists of nodes with outgoing edges leading into

v_i , nodes that the outgoing edges from v_i lead into, and nodes that have outgoing edges leading into the out-neighbors of v_i . For each node in $\{v_j\}$, the algorithm calculates the EI of a macroscale network after v_i and v_j are combined into a macronode, v_M , according to one of the macronode types in Figure 4. If the resulting network has a higher EI value, the algorithm stores this structural change and, if necessary, supplements the queue of nodes, $\{v_j\}$, with any new neighboring nodes from v_j 's Markov blanket that were not already in $\{v_j\}$. If a node, v_j , has already been combined into a macronode via a grouping with a previous node, v_i , then it will not be included in new queues, $\{v_j'\}$, of later nodes to check. The algorithm iteratively combines such pairs of nodes until every node, v_j , in every node, v_i 's Markov blanket, is tested.

Data Availability

All data used in this work were retrieved from the Konect Database [23] and also the Network Repository [24], which are publicly available. Software for calculating EI in networks and for finding causal emergence in networks is available by request or at <https://github.com/jkbren/einet>.

Disclosure

The opinions expressed in this publication are those of the author(s) and do not necessarily reflect the views of Templeton World Charity Foundation, Inc.

Conflicts of Interest

The authors declare no conflicts of interests.

Authors' Contributions

B.K. and E.H. conceived the project. B.K. and E.H. wrote the article. B.K. performed the analyses.

Acknowledgments

The authors thank Conor Heins, Harrison Hartle, and Alessandro Vespignani for their insights about notation and formalism of effective information. This research was supported by the Allen Discovery Center program through The Paul G. Allen Frontiers Group (12171). This publication was made possible through the support of a grant from Templeton World Charity Foundation, Inc. (TWCFG0273). This work was also supported in part by the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program.

Supplementary Materials

A: table of key terms. B: effective information calculation. C: deriving the effective information of common network structures. D: network motifs as causal relationships. E: table of network data. F: examples of consistent macronode. G: emergent subgraphs. (*Supplementary Materials*)

References

- [1] A.-L. Barabási, *Network Science*, Cambridge University Press, Cambridge, UK, 2016.
- [2] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, UK, 2010.
- [3] L. A. N. Amaral and J. M. Ottino, "Complex networks," *The European Physical Journal B—Condensed Matter*, vol. 38, no. 2, pp. 147–162, 2004.
- [4] A. Koseska and P. I. Bastiaens, "Cell signaling as a cognitive process," *The EMBO Journal*, vol. 36, no. 5, pp. 568–582, 2017.
- [5] F. A. Rodrigues, T. K. D. Peron, P. Ji, and J. Kurths, "The Kuramoto model in complex networks," *Physics Reports*, vol. 610, pp. 1–98, 2016.
- [6] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, "Activity driven modeling of time varying networks," *Scientific Reports*, vol. 2, p. 1, 2012.
- [7] D. S. Bassett and O. Sporns, "Network neuroscience," *Nature Neuroscience*, vol. 20, no. 3, pp. 353–364, 2017.
- [8] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: a network-based approach to human disease," *Nature Reviews Genetics*, vol. 12, no. 1, pp. 56–68, 2011.
- [9] N. Masuda, M. A. Porter, and R. Lambiotte, "Random walks and diffusion on networks," *Physics Reports*, vol. 716–717, pp. 1–58, 2017.
- [10] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [11] E. P. Hoel, L. Albantakis, and G. Tononi, "Quantifying causal emergence shows that macro can beat micro," *Proceedings of the National Academy of Sciences*, vol. 110, no. 49, pp. 19790–19795, 2013.
- [12] E. P. Hoel, "Agent above, atom below: how agents causally emerge from their underlying microphysics," in *Wandering towards a Goal: How Can Mindless Mathematical Laws Give Rise to Aims and Intention?*, A. Aguirre, B. Foster, and Z. Merali, Eds., pp. 63–76, Springer International Publishing, Berlin, Germany, 2018.
- [13] G. Tononi and O. Sporns, "Measuring information integration," *BMC Neuroscience*, vol. 4, no. 1, p. 31, 2003.
- [14] J. Pearl, *Causality*, Cambridge, New York, NY, USA, 2000.
- [15] J. Pearl, "Causal diagrams for empirical research," *Biometrika*, vol. 82, no. 4, pp. 669–688, 1995.
- [16] E. Hoel, "When the map is better than the territory," *Entropy*, vol. 19, no. 5, p. 188, 2017.
- [17] P. Erdős and A. Rényi, "On random graphs," *Publicationes Mathematicae*, vol. 6, p. 290, 1959.
- [18] B. Bollobas, "The evolution of random graphs," *Transactions of the American Mathematical Society*, vol. 286, no. 1, p. 257, 1984.
- [19] C. I. Del Genio, T. Gross, and K. E. Bassler, "All scale-free networks are sparse," *Physical Review Letters*, vol. 107, p. 1, 2011.
- [20] P. L. Krapivsky, S. Redner, and F. Leyvraz, "Connectivity of growing random networks," *Physical Review Letters*, vol. 85, no. 21, pp. 4629–4632, 2000.
- [21] A.-L. Barabási, R. Albert, and H. Jeong, "Mean-field theory for scale-free random networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 272, no. 1–2, pp. 173–187, 1999.
- [22] D. J. Watts and S. H. Strogatz, "Collective dynamics of "small-world" networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [23] J. Kunegis, "KONECT—the Koblenz network collection," in *Proceedings of the 22nd International Conference on World*

- Wide Web Companion*, p. 1343, Rio de Janeiro, Brazil, May 2013.
- [24] R. A. Rossi and N. K. Ahmed, "NetworkRepository: an interactive data repository with multi-scale visual analytics," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, p. 4292, Austin, TX, USA, January 2015.
 - [25] G. M. Edelman and J. A. Gally, "Degeneracy and complexity in biological systems," *Proceedings of the National Academy of Sciences*, vol. 98, no. 24, pp. 13763–13768, 2001.
 - [26] G. Tononi, O. Sporns, and G. M. Edelman, "Measures of degeneracy and redundancy in biological networks," *Proceedings of the National Academy of Sciences*, vol. 96, no. 6, pp. 3257–3262, 1999.
 - [27] P. K. Rubenstein, S. Weichwald, S. Bongers et al., "Causal consistency of structural equation models," 2017, <https://arxiv.org/abs/1707.00819>.
 - [28] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: bringing order to the web," Technical Report, Stanford Infolab, Stanford, CA, USA, 1998.
 - [29] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proceedings of the National Academy of Sciences*, vol. 105, no. 4, pp. 1118–1123, 2008.
 - [30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Hoboken, NJ, USA, 2012.
 - [31] J. Xu, T. L. Wickramaratne, and N. V. Chawla, "Representing higher-order dependencies in networks," *Science Advances*, vol. 2, Article ID e1600028, 2016.
 - [32] W. R. Ashby, *An Introduction to Cybernetics*, Chapman & Hall, London, UK, 1957.
 - [33] W. Marshall, L. Albantakis, and G. Tononi, "Black-boxing and cause-effect power," *PLoS Computational Biology*, vol. 14, p. 1, 2018.
 - [34] E. Laurence, N. Doyon, L. J. Dubé, and P. Desrosiers, "Spectral dimension reduction of complex dynamical networks," *Physical Review X*, vol. 9, Article ID 011042, 2019.
 - [35] D. D. Heckathorn and C. Cameron, "Network sampling: from snowball and multiplicity to respondent-driven sampling," *Annual Review of Sociology*, vol. 43, no. 1, pp. 101–119, 2017.
 - [36] G. Tononi, G. M. Edelman, and O. Sporns, "Complexity and coherency: integrating information in the brain," *Trends in Cognitive Sciences*, vol. 2, no. 12, pp. 474–484, 1998.
 - [37] A.-L. Barabási, E. Ravasz, and T. Vicsek, "Deterministic scale-free networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 299, no. 3-4, pp. 559–564, 2001.
 - [38] R. Griebenow, B. Klein, and E. Hoel, "Finding the right scale of a network: efficient identification of causal emergence through spectral clustering," 2019, <http://arxiv.org/abs/1908.07565>.
 - [39] A. R. Benson, D. F. Gleich, and J. Leskovec, "Higher-order organization of complex networks," *Science*, vol. 353, no. 6295, pp. 163–166, 2016.
 - [40] I. Scholtes, "When is a network a network? Multi-order graphical model selection in pathways and temporal networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p. 1037, Halifax, Canada, August 2017.
 - [41] H. Yin, A. R. Benson, and J. Leskovec, "Higher-order clustering in networks," *Physical Review E*, vol. 97, p. 1, 2018.
 - [42] R. Lambiotte, M. Rosvall, and I. Scholtes, "From networks to optimal higher-order models of complex systems," *Nature Physics*, vol. 15, no. 4, pp. 313–320, 2019.
 - [43] D. Gfeller and P. De Los Rios, "Spectral coarse graining of complex networks. Physical review letters," vol. 99, no. 3, p. 38701, 2007.
 - [44] S. Fortunato and D. Hric, "Community detection in networks: a user guide," *Physics Reports*, vol. 659, pp. 1–44, 2016.
 - [45] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, "Defining and identifying communities in networks," *Proceedings of the National Academy of Sciences*, vol. 101, no. 9, pp. 2658–2663, 2004.
 - [46] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, p. 1, 2004.
 - [47] K. Friston, "Life as we know it," *Journal of the Royal Society Interface*, vol. 10, Article ID 20130475, 2013.