




Article

Structural entropy constrains dynamics in directed networks

Blai Vidiella^{1,2} , Salva Duran-Nebreda³ , and Sergi Valverde^{3,4*} 

¹ ICREA-Complex Systems Lab, Universitat Pompeu Fabra, Barcelona

² Institut de Biologia Evolutiva, UPF-CSIC, Barcelona

³ Evolution of Technology Lab, Institute of Evolutionary Biology (CSIC-UPF). Pg. Barceloneta 37, 08003 Barcelona.

⁴ European Centre for Living Technology (ECLT), San Marco 2940, 30124, Venezia, Italy.

* Correspondence: sergi.valverde@ibe.upf-csic.es

Abstract: Complex systems can be described as the network of interactions between their components. In directed networks, information, energy and matter flow in dominant directions. Undirected networks, on the other hand, cannot easily capture these asymmetries. The local flow of information depends on the node spin, which is a function of the number of incoming and outgoing links attached to it. And yet, how the heterogeneous association between in-degree and out-degree constraints global dynamics remains poorly understood. Here, we develop a new theoretical model to study the relationship between node directionality and global flow dynamics. Our computational analysis of random and scale-free directed networks indicates that structural diversity constrains the dimensionality of reachable dynamical space. These results suggest how local directionality might be an universal driver of global dynamics in many systems, from social networks, to technological systems to the connectome.

Keywords: Structural Entropy; Dimensionality; Flow Dynamics; Spin; Directed Network; Scale-free

1. Introduction

Underlying many systems, we can find a network of interactions. Many different studies have characterised their properties and modelled their evolution [1]. Undirected graphs are one of the most common approaches. These are very useful when looking at reciprocal interactions, but they are less adequate when describing asymmetric interactions between elements. In natural and artificial systems, information, energy and matter flow in dominant directions. Mapping this flow will uncover their general design principles [2]. A classic example is a colony of social insects, where information scales with mobility and group size [3]. The common trait in insect and animal communities is the establishment of hierarchical structures. These networks canalize the flow of information. For example, when two individuals meet in paper wasp communities, they adopt either a dominant or submissive role, creating a linear hierarchy across the colony [4]. Hierarchies are also a fundamental component in the evolution of cultural and technological systems. Discontinuities and sudden technological change depend on the cost of storing and (reliably) communicating large amounts of information [5,6]. Again, a massive network connecting people, institutions and cities, works as the informational backbone enabling social cooperation and information transmission. In all these cases, directed networks emerge as the natural approach to study the transmission of diseases, resources and innovations through asymmetrical structures.

What are the general principles and constraints governing the structure-dynamics relationship? Complexity emerges from simpler systems by means of growth and evolutionary processes, which in turn, are shaped by physical constraints. In information networks, specific node properties (like the rate of processing) influence how information is sent from one node to its neighbors. This is the case for traffic congestion, which

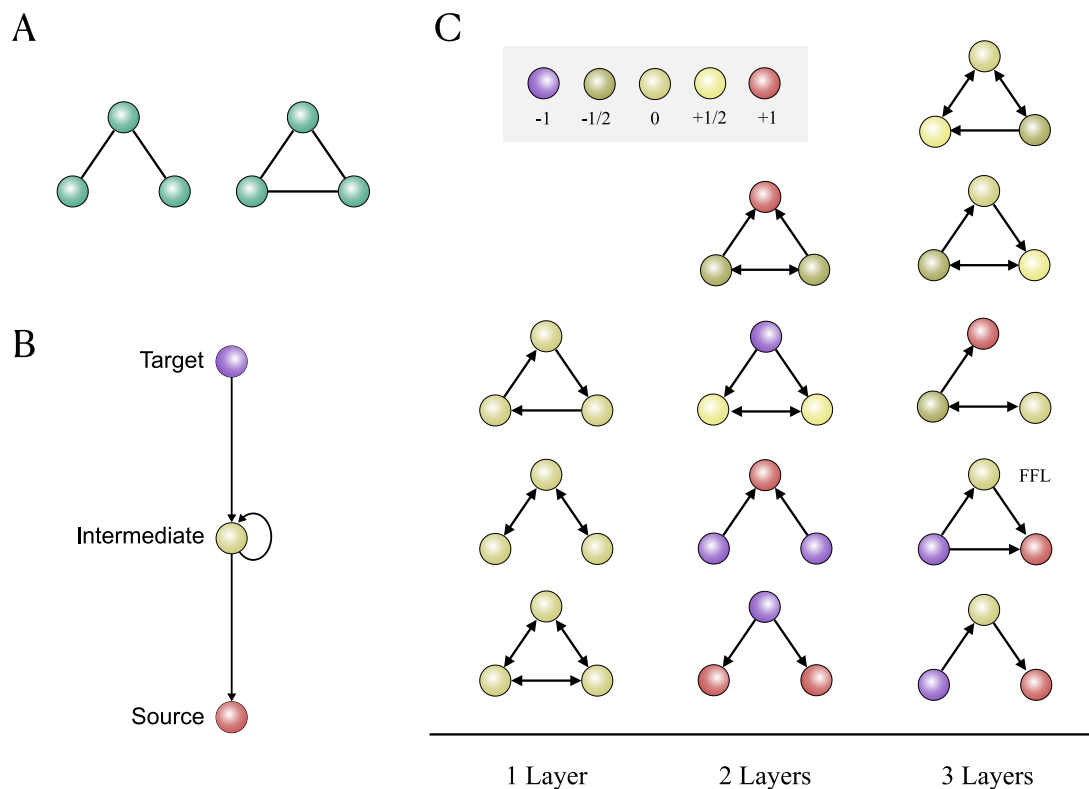


Figure 1. Undirected vs directed networks. (A) The set of undirected networks with $N = 3$ nodes consists of only two topologies: chains and triangles. On the other hand, nodes in a directed network belong to different classes or "spins" $s \in S$ (see text). (B) Target nodes (blue) have $K_1 = 0$, intermediate nodes (yellow) have both types of links ($K_1 > 0$ and $K_2 > 0$) and source nodes (red) only have incoming links ($K_2 = 0$). (C) By using directed links, we can expand the universe of undirected networks to 12 different networks using a collection of $|S| = n = 5$ different spins (see inset).

seems to be an inevitable result of user's behaviour coupled to network dynamics. In this context, theoretical models suggest that routing algorithms can take advantage of the Internet's network topology to maximize global performance [7].

In this paper, we present a simple model mapping the relationship between structure and dynamics. This model is inspired by observations of flow dynamics in natural and artificial systems (including studies of computational systems [8]). We assume that structural patterns are the main driver in the dynamics of information, and depends much less on system-specific traits. This is obviously a simplification, but we hope to uncover general principles in the structure-dynamics relationships. We simulate our model on different topologies and quantify their impact on flow dynamics. A main difference between our study and previous approaches (e.g., [9]) is the in-degree and out-degree sequences are not considered separately. Instead, we focus on the spin sequence, which conflates in-degree and out-degree in a single structural trait [10]. Spins have been interpreted as the functional role played by the nodes, because they restraint the rate of information going through them. Our results show that, at the global level, the diversity of spins constraints the average activity in the system, and also the dimensionality of the space that can be explored by network dynamics. The implications for the study of directionality on the evolution of complex systems are outlined.

2. Structure

2.1. Node Spin

We start by defining structural measures of node diversity (to be compared with dynamical properties). Structural diversity depends on the possible ways we can connect one node to each other. For example, the space of (connected) undirected networks with

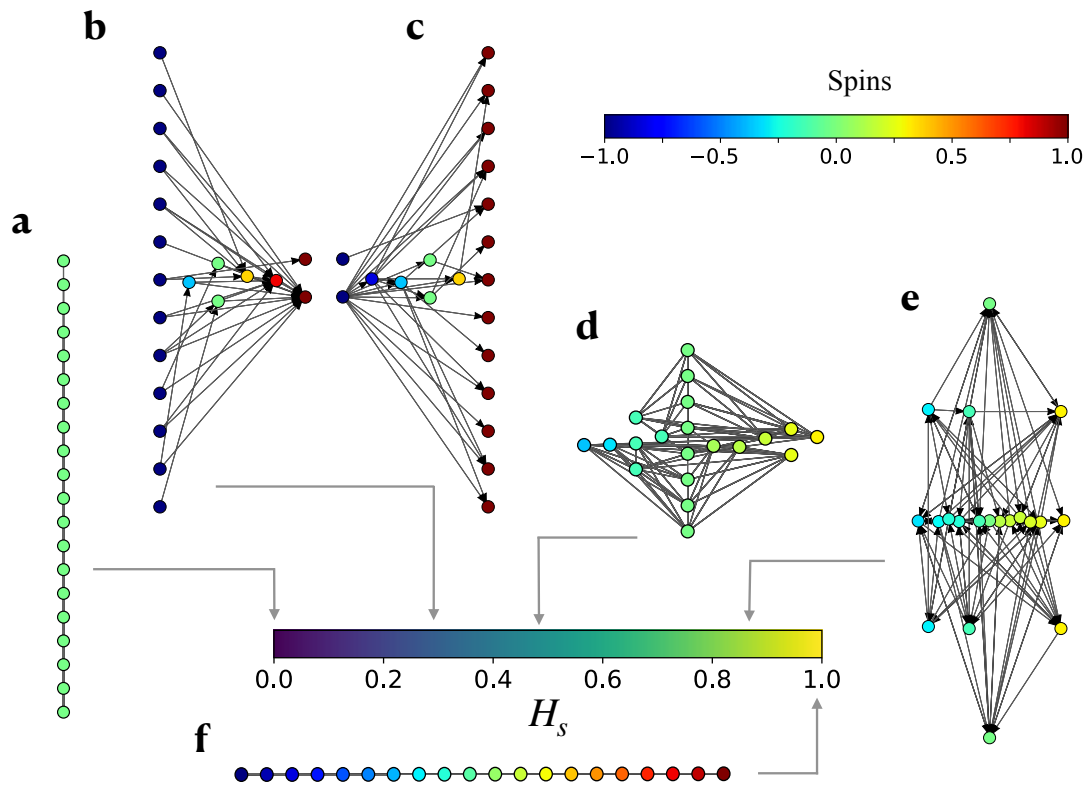


Figure 2. Location of directed networks in the structural entropy ($H / \log N$) space. Every node in the (a) minimal entropy network ($H = 0$) displays the exact same spin, $P(s_i = 0) = 1$. Notice how multiple systems can be mapped here, including the loop. Between the two extremes of this space, we can find all possible directed networks, including: (b) the GNC network and its (c) transposed (see text), (d) random modular networks, and (e) random directed networks. Finally, (f) entropy is upper-bounded ($H \leq \log N$) by a network with a uniform distribution of spins, $P(s_i) = 1/n$ (see text).

$N = 3$ nodes consist of only two possible designs: chains and triangles (see Figure 1A). The same space is far richer in directed networks because we can combine nodes in many different ways (see Figure 1C).

A directed network $G = (V, E)$ can be represented with the (possibly asymmetric) adjacency matrix $A = [A_{ij}]$ where $A_{ij} = 1$ if node i interacts with node j and $A_{ij} = 0$ otherwise. In an undirected graph, K_i denotes the degree of a node, or the total number of node interactions. Because our networks are directed, there are two possible definitions of node connectivity. Let's define the in-degree

$$K_1(i) = \sum_j A_{ji}$$

which gives the total number of incoming links, and the out-degree

$$K_2(i) = \sum_j A_{ij}$$

which gives the number of links out-going from any node.

It has been proposed that the functional role of nodes depends on the correlation between in-degree and out-degree. For example, a microchip is composed by a (potentially very large) set of interconnected logical gates. Here, we can interpret $K_1(i)$ as the number of inputs and $K_2(i)$ as the number of outputs attached to a gate. Balance between inputs

and outputs will surely constrain functionality. For example a multiplexor [11] is a device that produces a single output from multiple inputs. On the other hand, complex circuits with many inputs and outputs can be made by combining several functions into a single device. By looking at the interplay between in-degree and out-degree, we can derive the following classification of node types:

1. **Sink** nodes define elementary units. Sink nodes can be thought as the atoms of the network, that is, nodes that do not depend on any other element, i.e., their out-degree is $K_2(i) = 0$.
2. **Target** node relies upon a larger number of outgoing connections, but no incoming links, i.e., their in-degree $K_1(i) = 0$.
3. **Intermediate** node takes many incoming links (i.e., the in-degree is at least one) and generates several outgoing connections (i.e., the out-degree is non-zero). Many directed networks are composed by intermediate parts acting as the "glue" connecting targets to the sink nodes. A particular case is an intermediate node with exactly the same number of incoming and outgoing links, i.e., $K_1(i) - K_2(i) = 0$.

The above can be properly quantified with the concept of "spin" (see Fig. 1B). The spin $-1 \leq s(i) \leq 1$ of a node is the difference between their (normalized) in-degree and out-degree:

$$s(i) = \frac{K_1(i) - K_2(i)}{K(i)} \quad (1)$$

where $i = 1, \dots, n$ [10]. We can see that the limit cases for the spin value correspond to the source ($s(i) = +1$) and target ($s(i) = -1$) nodes. Figure 1C classifies small directed networks depending whenever the set of network nodes displays one (left), two (middle) or three (right) spin types.

2.2. Structural Entropy

For each directed network, we can obtain a finite set \mathcal{S} of $s(i)$ of spins. We aggregate this local index to derive a global diversity measure. One way is using Shannon's entropy associated with the spin distribution $P(s)$. Let's define structural entropy of the network as the entropy of spins:

$$H(s) = - \sum_{s \in \mathcal{S}} P(s) \log P(s) \quad (2)$$

where $P(s)$ is the distribution of nodes having spin s . Larger values of structural entropy $H(s)$ reveal a multiplicity of spin values $s(i)$, whereas networks displaying a few (or no) intermediate nodes have low entropy. Structural entropy is a robust measure of network diversity, which is closely linked to the presence of multiple layers of complexity. Figure 2 displays several networks and its location along the $H/\log N$ axis. We can see the network layout maps spin values to vertical layers. Nodes in the network have been arranged so leftmost and rightmost layers correspond to target ($s_i = -1$) and sink ($s_i = 1$) nodes, respectively, while all intermediate nodes with $s_i = 0$ are placed at the center.

A lower bound for entropy is given by the network composed only with n_s source and n_t target nodes, without any intermediate nodes. We have $P(s = -1) = P(-1) = n_t/N$ and $P(s = 1) = P(1) = n_s/N = 1 - P(-1)$ where $N = n_s + n_t$. We can easily compute the structural entropy as

$$H_m(1) = -P(1) \log P(1) - [1 - P(1)] \log [1 - P(1)]$$

. This would be zero at the two limits and its symmetric case. As soon as we increase the number of intermediate nodes, the entropy will increase. Another limit case would be

the maximally diverse network with n different spin values having the same probability $P(s)$, i. e. $P(s) = 1/|\mathcal{S}|$ for all $s \in \mathcal{S}$, which would give:

$$H_n = - \sum_{s \in \mathcal{S}} \left(\frac{1}{n} \right) \log \left(\frac{1}{n} \right) = \log n$$

. Then, the quantity $H/\log(n)$ represents the normalized entropy of spins, which can be used when comparing different types of networks.

These extremes represent limit cases for real-world systems (see Figure 2). A directed network with maximal entropy could be interpreted as an idealized representation of linear dominance hierarchies. On the other hand, a loop has the minimal entropy because all nodes have exactly the same spin. Notice that there are multiple networks with $H = 0$ (e.g., undirected networks).

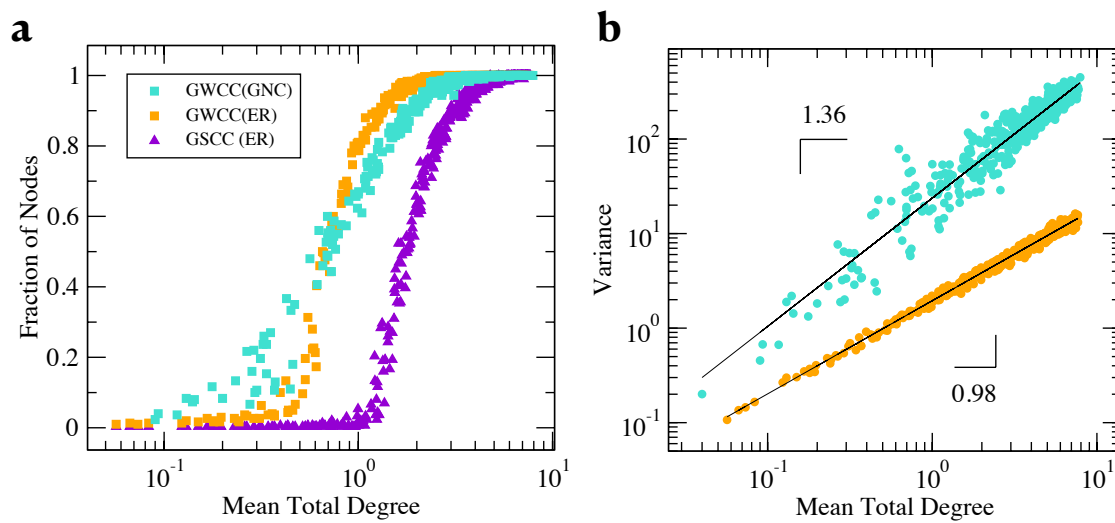


Figure 3. Structural measures of random and GNC networks as a function of their average degree $\langle K \rangle$: (a) fraction of nodes in the giant weekly connected component (GWCC) and giant strongly connected component (GSCC), in the ensemble of random (orange and violet, respectively) and GNC (turquoise) networks. (b) Degree variance scales with average degree, $\sigma_K^2 \sim \langle K \rangle^\alpha$, with $\alpha = 0.98$ for random networks and $\alpha = 1.36$ in GNC networks. Every point corresponds to one of 800 simulated networks with $N = 300$ nodes each (see text).

2.3. Network models

To determine the influence of structural heterogeneity in global dynamics, we will compare random and scale-free networks. A random directed graph is an example of static, disordered structure. Over the years, the study of complex networks has been strongly influenced by theoretical developments in random graph models. Random graphs, although very simple, provide a basis for more sophisticated analyses. Many studies, e.g., the classic Poisson random graph defined by Erdős and Rényi [12,13], focus on undirected graphs, but more recently there has been a surge of interest in the analysis of directed random graphs [14].

Definition of the random directed graph $G_{rnd} = (V, E)$ is straightforward. Given a set of N nodes and a probability $p \geq 0$, we set a directed link $(i, j) \in E$ with probability p . It can be shown that both the in-degree and out-degree distributions $P(K_i)$ ($i = 1, 2$) follow the Binomial distribution

$$P(K_i) = \binom{N}{K_i} p^{K_i} (1-p)^{N-K_i} \quad (3)$$

with parameters $(N - 1, p)$ and expected total degree $\langle K \rangle = (N - 1)p$. Notice that the average in-degree and the average out-degree are the same because $\sum_i K_1(i) = \sum_j K_2(j) = m$, where m is the total number of links.

We will compare random graphs with a growing network model [15]. This model reproduces the statistical features of acyclic directed networks, including the World-Wide Web [15], patent citation networks [16], software networks [17], and the evolution of programming languages [18]. The acyclic nature of these networks arises as a consequence of some kind of node ordering. Unlike random graph models, nodes in a growing network do not receive new links independently from others (see [19]). Instead, attachment probability depends on intrinsic node properties, e.g., ranking in a social group [20] or node importance [21]. Either because temporal dependencies or other constraints, the observed structure in real systems often departs considerably from random graphs.

In the model of growth by copying (GNC), a network $G_{gnc} = (V, E)$ grows by adding one node at each time step. This new node $v_i \in V$ links to m randomly chosen source nodes $v_j \in V$ with probability p , as well to all (ancestor) nodes $v_k \in V$ pointed by the source $(v_j, v_k) \in E$, with probability q . Notice that mp and mq are the average number of source and ancestor nodes, respectively. The in-degree distribution is [15]:

$$P(K_1) = \frac{N}{(K_1 + 1)(K_1 + 2)} \quad (4)$$

for $i < N - 1$, while $P(K_1) = 0$ for $i \geq N$. Asymptotically, the exponent of the in-degree distribution $P(K_1) \sim K_1^{-2}$ is independent of parameters, including the attachment probabilities p and q . On the other hand, the predicted out-degree distribution has Poisson form. Krapivsky and Redner also computed the expression for the joint degree distribution $N(K_1, K_2)$ or average number of nodes with in-degree K_1 and out-degree K_2 . For a very large network ($N \rightarrow \infty$), the leading behaviour of its joint distribution is: [15]:

$$N(K_1, K_2) \rightarrow \left[\frac{(\log N)^{K_2-1}}{(K_2 - 1)!} \right]^{K_1+1} \quad (5)$$

We also consider transposed (or reversed) networks (see Figure 3c). Given a directed network G , the transposed network G' is the same graph with the direction of every link reversed. We can easily compute the transposed graph as follows. If $A = [A_{ij}]$ is the adjacency matrix of G , the transposed graph has adjacency matrix $A^T = [A_{ji}]$.

2.4. Structure of Directed Networks

In this paper, we study the relationship between structure and dynamics using an empirical ensemble of random and GNC networks. Here, we characterise the main structural properties of these networks (their dynamical properties are determined below). To show this, we have sampled 400 networks (for each model) at random from the range $0 < \langle K \rangle < 8$ of small degrees, which is typical of real systems. To ensure good statistics, networks of $N = 300$ nodes are generated. We sampled each network using a random combination of the model parameters. For the GNC network, we fixed $mq = 1$ and sampled the attachment probability p .

Average connectivity $\langle k \rangle$ is the most important index for any network, and imposes strong constraints on structural diversity. Connectivity also defines the limits for any dynamical processes taking place on top of networks, like transport or the diffusion of information. In order to preserve their functions, a network must be connected, i.e., there must be a path connecting any pair of nodes. An important theoretical prediction is the emergence of connected paths happens very rapidly once a critical threshold of connectivity is crossed.

Directed networks experience two types of percolation transitions associated to the (so-called) giant components [22]. The first transition maps the emergence of the giant weakly connected component (GWCC), the fraction of nodes that are reachable from every other node in the network when link direction is ignored. This is followed

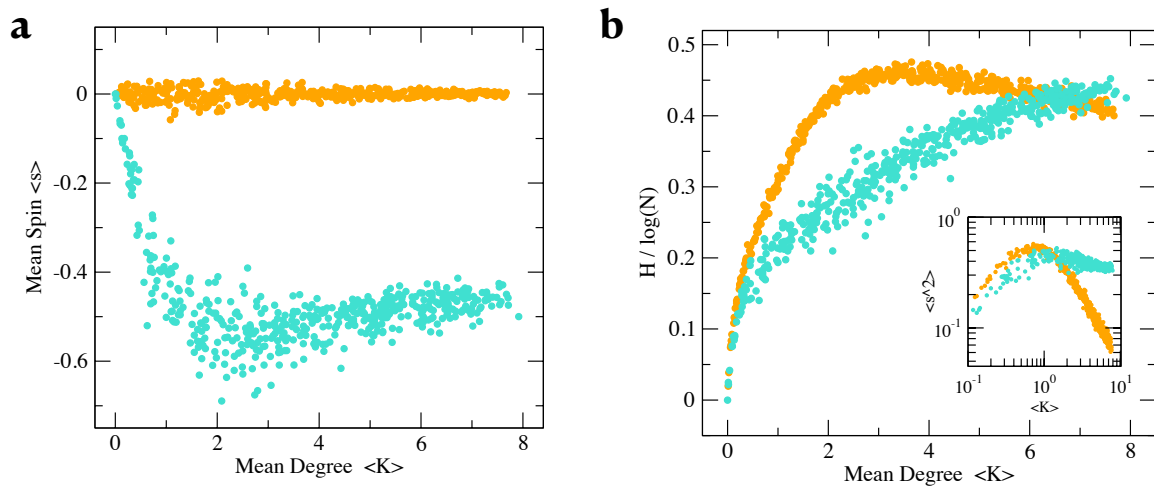


Figure 4. Dependence of (a) average spin $\langle s \rangle$ with mean degree $\langle K \rangle$ in random (orange) and GNC (turquoise) networks. Random networks display an abundance of intermediate nodes ($\langle s_i \rangle \approx 0$) while GNC networks are clearly biased towards negative spin. (b) Normalized entropy $H / \log(N)$ vs $\langle K \rangle$ in GNC (turquoise) and random (orange) networks. Entropy in random networks saturates once the GSCC emerges, which is consistent with an exponential decay of spin variance σ_s^2 (see inset). To obtain an accurate entropy estimate, we use histograms of binned spins (20 bins in the interval $(-1, 1)$). In these plots, every point corresponds to one simulated network with $N = 300$ nodes (see text).

by another transition towards the giant strongly connected component (GSCC). In this second transition, all pairs of nodes are connected by a directed path (all links along this path are aligned). As we will see, the size of the GSCC is a key parameter for the network dynamics. Figure 3(a) shows the location of percolation transitions in our ensemble of random and GNC networks. Exact location of percolation thresholds can be calculated using the maximum eigenvalue of the adjacency matrix [23].

A main difference between random and scale-free networks is the greater degree heterogeneity of the latter. In this context, one of the most widespread and intuitive measures of heterogeneity is the degree variance $\sigma_K^2 = \langle K^2 \rangle - \langle K \rangle^2 = (1/N) \sum_i K(i)^2 - \langle K \rangle^2$ ([24]). Figure 3(b) shows the scaling of degree variance $\sigma_K^2 \sim K^\beta$ in both the random ($\beta \approx 1$) and GNC ($\beta \approx 1.36$) networks. For the random graph, it is easy to show that $\langle K \rangle = Np$ and $\sigma_K^2 = Np(1-p)$, that is, $\sigma_K^2 \sim \langle K \rangle$. On the other hand, the super-linear scaling observed in the GNC network derives from the slow decay of fluctuations in the probability distribution of the total number of links [15]. In general, dependence of degree variance with density and size requires a proper measure of normalisation [25].

The diversity of spins is an useful index for directed networks. Random networks are dominated by intermediate nodes ($s_i = 0$), while GNC networks are biased towards negative spins, i.e., target nodes. Figure 4(a) shows the average spin $\langle s \rangle \approx 0$ is roughly constant for random graphs with independence of $\langle K \rangle$. For GNC networks, average spin is negative and saturates (once the GWCC emerges) around $\langle s \rangle \approx -0.5$. This asymmetry is more pronounced in denser networks. As shown in Figure 4(b) the structural entropy of random networks saturates once the GSCC is reached. This is consistent with the inverse dependence between structural entropy and the variance of spins σ_s^2 in random directed graphs. On the other hand, we observe a scaling regime of entropy in the subset of GNC networks, which is associated with the maintenance of spin diversity.

3. Dynamics

3.1. Flow networks

We now develop our abstract model for the relationship between structure and dynamics. This model borrows inspiration in natural and artificial systems, including neural networks, electronic circuits and communication networks. Here, we do not

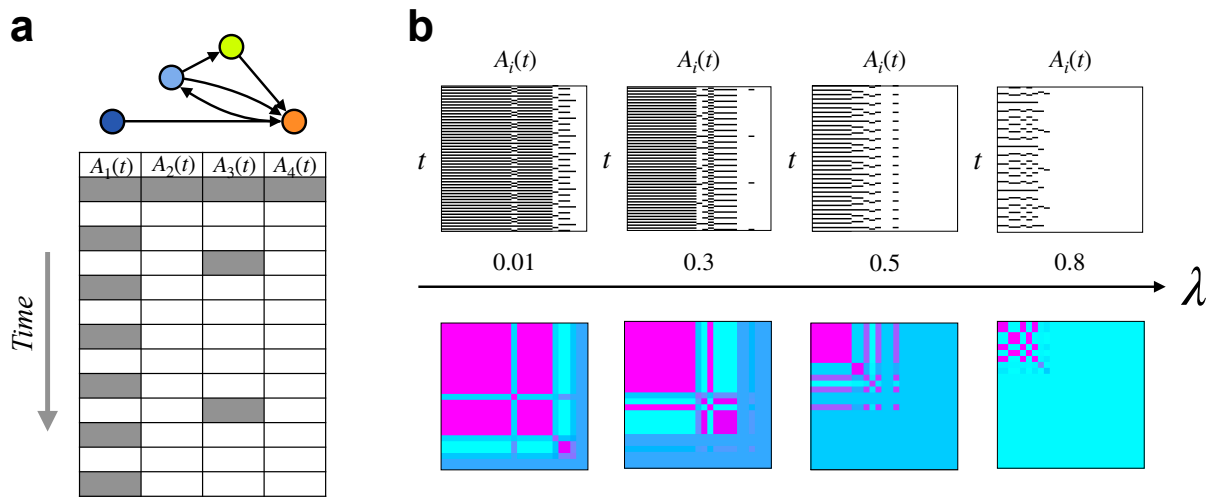


Figure 5. Spatio-temporal dynamics in flow networks. (a) Activity in this small network generates the pattern below ($\lambda = 0.1$). Each column describes the activity of the node $i \in [1, 4]$, where $A_i(t) = 1$ (white) or $A_i(t) = 0$ (gray). Nodes have been sorted in ascending spin order and time runs downwards. Notice how structural coupling (e.g., bi-directional links) can induce activity correlations, i.e., between $A_2(t)$ and $A_3(t)$. (b) Spatio-temporal dynamics in a GNC network with $N = 25$ nodes (top). From left to right, we can see the evolution of activity $A_i(t)$ at each node $1 \leq i \leq N$, when input flow varies from $\lambda = 0.01$ to $\lambda = 0.8$. For each spatio-temporal diagram, we compute the corresponding covariance $C = [C_{ij}]$ matrix (bottom) used to estimate the dimensionality of flow dynamics (see text). (c) Covariance matrix measures activity correlations between pairs of nodes, which are partly determined by the global input load λ (see text).

include specific functions of any system, like threshold functions in neuronal networks (cite) or logical functions implemented in an electronic circuit [8]. Instead, we simply assume that the only function performed by the nodes is to receive and forward information to its neighbouring nodes. By focusing on the dynamics of communication alone, we can observe the general relationship between directionality and global dynamics in absence of system-specific biases (e.g., different functions and/or input signals).

Let us consider a network $G = (V, E)$, where nodes $v_i \in V$ can store and exchange information with their neighbors. This network is an abstract model of a real system where information, energy or matter flows through links. Nodes in this *flow network* receive information from incoming links, which is then forwarded only through the outgoing links. For simplicity, we assume that nodes have infinite storage capacity, but they can redirect only a finite amount of their current load. Let's define the load $B_i \geq 0$ as the amount of information stored at the node. At each time step, we apply the following set of rules:

1. Initial condition: All nodes are empty at the beginning, $B_i(0) = 0$.
2. Input flow: A fraction $\rho > 0$ of nodes increase their load by an amount $0 \leq \lambda \leq 1$.
3. Flow of information: At each time step t , every v_i node decreases its load, $B_i(t) = B_i(t-1) - w$, by a fixed amount $w \sim 1/K_2(i)$, and forwards the extracted information through each outgoing link $(i, j) \in E$, that is, the receiver node v_j increases its load by $B_j(t) = B_j(t-1) + w$.
4. Activity: during each timestep each node is evaluated in terms of load, a node v_i is considered active if its load $B_i > 0$.

Figure 5A shows the results of our dynamical flow model on an example small graph. For each node v_i we show its dynamical activation state, i. e. whether its load $B_i > 0$. As time progresses, specific cyclical patterns of activation are established, which are periodically repeated. These patterns of activation and inactivation depend on the magnitude of input flow λ (Figure 5B top), as λ increases and approaches 1, the system becomes saturated and most nodes are always active. The spatiotemporal diagram can

also be analyzed with the covariance matrix (Figure 5B bottom). This method allows us to quantify node activity correlations that reduce the dimensionality of the system dynamics (see below).

3.2. Mean-field model

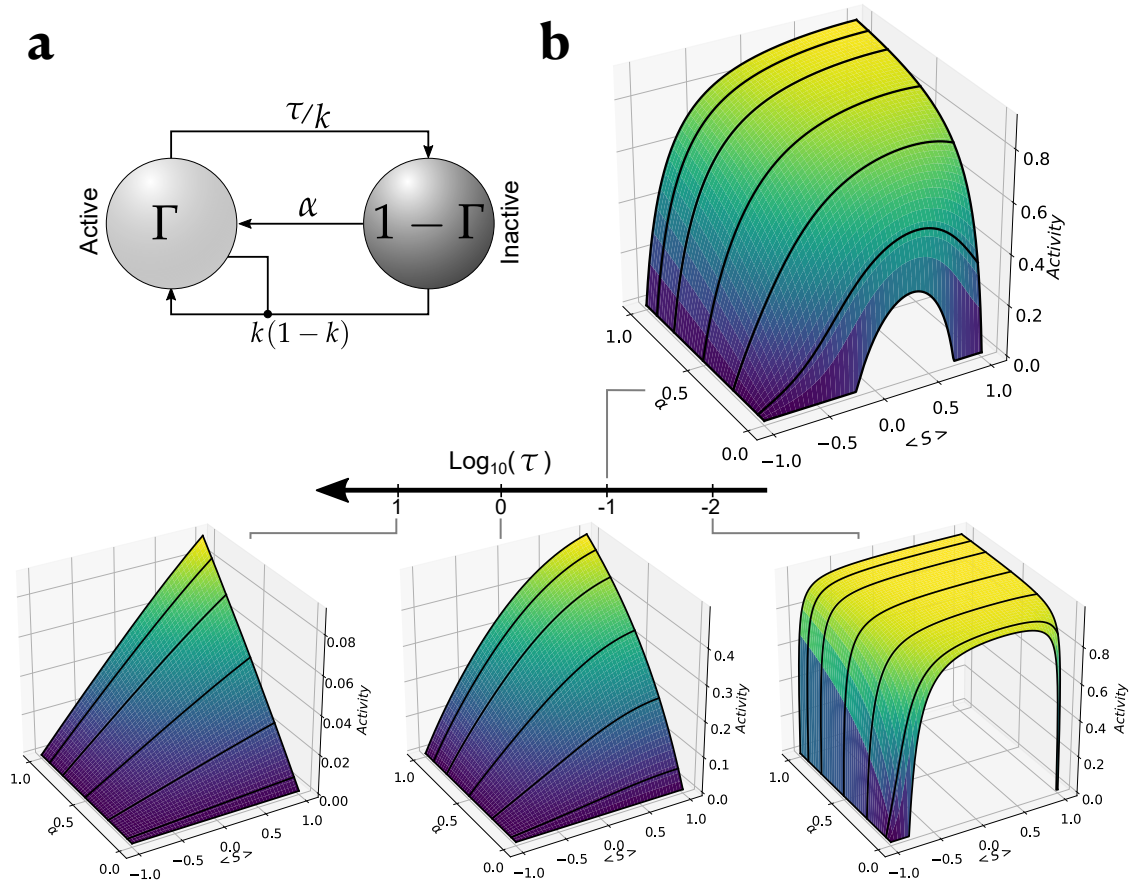


Figure 6. Mean-field model predicts differences in average activity depending on mean spin ($\langle s \rangle$) and activity life-time (τ). (a) Schematic representation of the mean-field approach to temporal evolution of activity Γ in flow networks (see text). (b) Mean-field prediction for average activity $\langle \Gamma \rangle$ as a function of mean spin $\langle s \rangle$ and activation rate α . Each surface corresponds to a different average lifetime (τ^{-1}): $\tau = 0.01, \tau = 0.1, \tau = 1$ and $\tau = 10$.

Here, we develop a mean-field approximation for the long-term evolution of the average activity $\langle A(t) \rangle$ of the flow network. These population models have been frequently used when describing the dynamics of neuronal networks [26,27]. Mean-field models offer a simpler, qualitative picture of the dynamics on networks. Moreover, their predictions can be used whenever empirical observations of average activity are available.

Instead of looking at the average load in the system, we here consider a simpler approach focusing on average node activity

$$\Gamma = \frac{1}{N} \sum_{i=1}^N \Gamma_i, \quad (6)$$

where the threshold function $\Gamma_i = \delta(B_i > 0)$ indicates if the node is active ($\Gamma_i = 1$) or not ($\Gamma_i = 0$). We assume that nodes can be activated by their own load or the dynamics of the neighbouring nodes. On the other hand, activity decreases when a fraction of node is released to the neighbors. Both processes are mediated by local connectivity (see below).

The time evolution of Γ is defined as follows (see schema in Figure 6):

$$\frac{d}{dt}\Gamma = (1 - \Gamma)[k(1 - k)\Gamma + \alpha] - \frac{\tau}{k}\Gamma$$

where $k \equiv K_2/K$ is the probability of outgoing link, $K = K_1 + K_2$ is the total (undirected) degree, τ^{-1} is the average lifetime of information, and α is the autonomous activation rate. Notice that $K_1/K + K_2/K = 1$ and thus, $1 - k$ defines the fraction of incoming links. The number of active nodes increases when three conditions are satisfied: there must be both inactive ($(1 - \Gamma) > 0$) and active ($\Gamma > 0$) nodes in the network, probability of a link between two nodes is $k(k - 1) > 0$, and/or the activation rate is $\alpha > 0$. Nodes are deactivated at a rate $(\tau/k)\Gamma$, which is a function of the average time need to escape the system and inversely related to the average out-degree $1/k$ (see below).

This system is not defined for $k = 0$ because the term $1/k \rightarrow \infty$. On the other hand, the long-term activity when $k = 1$ is $\Gamma = \alpha/(\alpha + \tau)$. For the other values of the out-degree k , the system has the following fixed points:

$$\Gamma_{\pm} = \frac{k^2(1 - k) - \alpha k - \tau \pm \sqrt{(k^2(1 - k) - \alpha k - \tau)^2 + 4\alpha k^3(1 - k)}}{2k^2(1 - k)} \quad (7)$$

These fixed points will exist when the square root is positive, that is, when $\alpha > \alpha_c$. The critical value for α is

$$\alpha_c = -\left(k(1 - k) + \frac{\tau}{k}\right) \pm 2\sqrt{\tau(1 - k)}. \quad (8)$$

Given that $0 \leq k \leq 1$, the above condition is always achieved for $\alpha \geq 0$. When no external load is received ($\alpha = 0$), the system can relax to two fixed points. One is the complete extinction of the activity ($\Gamma_0^* = 0$) or a remnant $\Gamma^* = 1 - \tau/(k^2(1 - k))$. The stability of these fixed points depends on

$$\frac{d}{d\gamma} \frac{d\Gamma}{dt} = (1 - 2k)k(1 - k) - \tau/k.$$

Extinction of activity is stable if $\tau > k^2(1 - k)$.

Figure 6 shows the fixed point for the activity of nodes Γ with mean spin $s = \langle s_i \rangle = 2k - 1$, activation rate α , and average lifetime τ^{-1} . Our flow model suggests the bias towards negative average spin $k < 0.5$ deactivates nodes faster than an abundance of positive or neutral spins. Activity can also be increased by processing delays (τ). Panels in figure 6b show the stable fixed point for different parameter's combinations. If activation is quickly lost ($\tau^{-1} \rightarrow 0$), the system will tend to be in a low activity condition, and practically independent from network structure ($\langle s \rangle$). On the other hand, the network is rapidly saturated when activity is well preserved ($\tau^{-1} \rightarrow \infty$) (see Panel $\tau = 10^{-2}$). Between these two cases, the interplay between the structure ($\langle s \rangle$) and self-activation rate (α) defines the system's activity state. Depending on the average lifetime, low α minimize the structural effects (see Panel $\tau = 1$). Exceeding activation rates ($\alpha \approx 1$) always saturates the network, for a wide range of structures. However, and independently of α and τ , low activity levels are always associated with small spin averages ($\langle s \rangle < 0$).

3.3. Entropy and Dimensionality

As we have seen, a directed network consists of many different elements. Link directions can generate an heterogeneous distribution of flow, where distinct regions of the system are subject to different loads. Averaging such collective dynamics cannot, in general, accommodate for this structural variance. A concise representation of complex activity necessitates a large number of variables, perhaps as many as the number N of

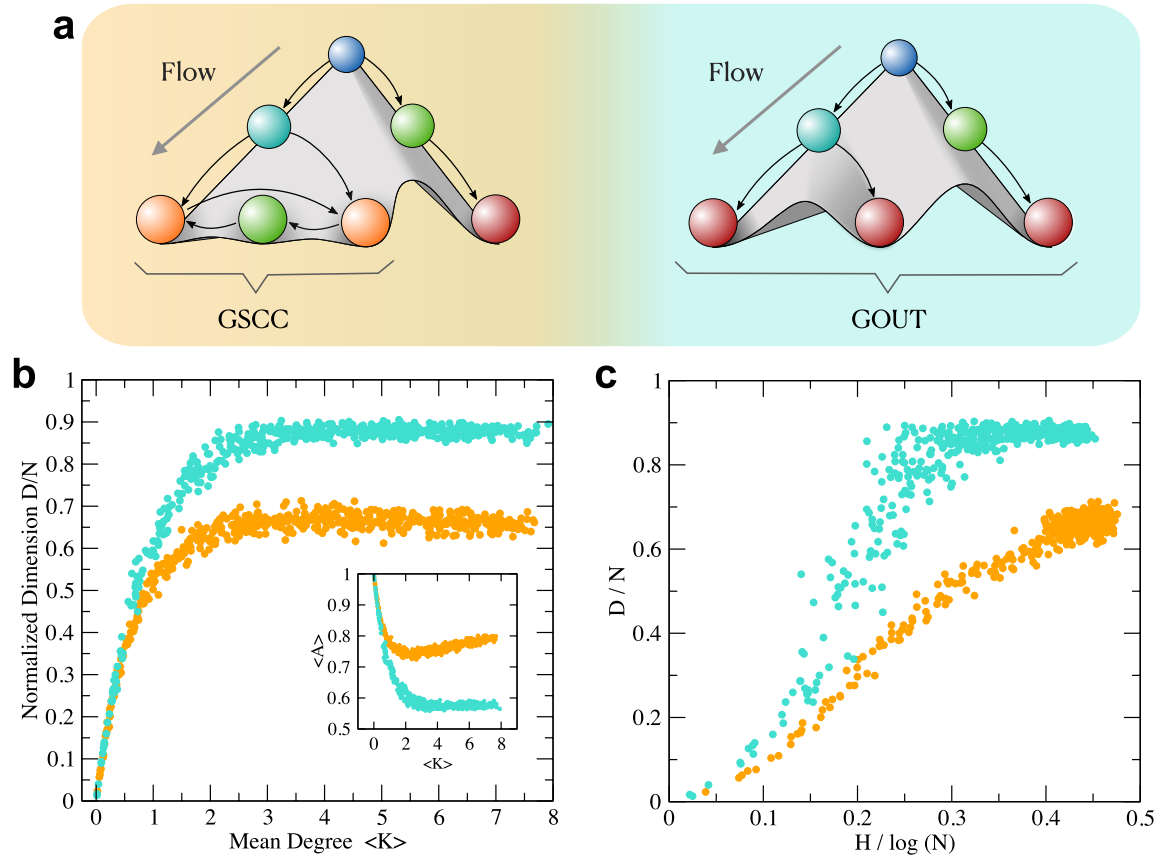


Figure 7. Relationship between structure and dynamics in flow networks. (a) Illustration showing how activity is characteristically feed-forward (it flows towards the GOUT component, see right), but it can also be trapped by the GSCC component of cyclic networks (left). 3-D visualizations of empirical networks can be seen at this [website](#). (b) Normalized dimensionality D/N in GNC networks (blue) is larger than in random networks (orange) due to the low activity of the former. Inset: Our simulations shows that (b) mean activity is larger in random networks (orange) than in GNC networks (blue) due to the large GSCC component of random graphs (see Figure 3A). On the other hand, nodes in GNC networks are significantly less active due to the acyclic nature of the network. A large number of dimensions corresponds to a more flexible dynamical system. (c) Correlation between normalized dimensionality and structural entropy in random (orange) and GNC (turquoise) networks. Dimensionality is bounded by the interplay between structural diversity and mean activity (see text). Each point in the plots corresponds to a simulated network from our ensemble (see text). In all simulations, the input flow is $\lambda = 10^{-4}$.

nodes in the network (but see [28]). In this context, dimensionality provides a compact measure of the degree of coordination between all pairs of nodes. For example, it has been suggested that high-dimensional activity in neural networks reduces dynamical constraints and enables flexible computations [29,30].

Dimensionality D is a (weighted) measure of the number of degrees of freedom explored by network dynamics [31]:

$$D = \frac{(\text{Tr } C)^2}{\text{Tr } C^2} = \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2} \quad (9)$$

where $C = [C_{ij}]$ is the co-variance matrix of the vectors of node activity $\mathbf{A}(t) = [A_i(t)]$, with $1 \leq t \leq T$, $\text{Tr } C = \sum_{i=1}^N C_{ii}$ is the trace of the covariance matrix, $\text{Tr } C^2 = \sum_{i=1}^N C_{ij}C_{ji}$, and λ_i is the i -th eigenvalue of the covariance matrix. The eigenvector of the covariance matrix C are the axes spanning the space filled by the cloud of activity vectors. If the components of $\mathbf{A}(t)$ are independent and have equal variance, all eigenvalues have the

same value and $D = N$. On the other hand, when the activities of nodes are highly correlated, the variance evenly spreads across a smaller number $D < N$ of dimensions.

A recent theoretical study predicts the dimensionality of spiking recurrent networks from their local structural features and the type and magnitude of input received. They concluded that the number of dimensions strongly correlates with the network connectivity [32]. Here, we perform simulation experiments to test if a similar structure-dynamics relationship holds in flow networks. The dimensionality of spiking networks tends to be low and it is regulated by structural motifs, i.e., small subgraphs with non-random frequencies [32]. Our structural features are not network motifs, but the distribution of spins. We conjecture that spins play a role in the dynamics of directed networks. Interestingly, [32] has shown that reciprocal motifs had the biggest contribution to network dimensionality. Bi-directional links are related to reciprocal motifs, which can be further associated with intermediate nodes (those nodes having $s \approx 0$ spin).

In our experiments, we have simulated the temporal evolution of flow networks in our ensemble, and for $t_0 = 10N$ iterations. After this transient, we gathered vectors of node activity $A_i(t)$ during the interval $t_0 \leq t \leq (t_0 + T)$, where $T = 64$ time steps. Using the two sets of activity vectors (one for each type of network), we computed the covariance matrices C_i and their dimensionality D_i ($1 \leq i \leq 400$).

Our simulations of flow dynamics show high or low dimensionality D depending on the interplay between input rate λ and the structural diversity of the network. As suggested by the mean-field approach, saturation of mean activity yields highly correlated states in both random and GNC networks and a minimal dimensionality ($D \rightarrow 0$). This is consistent with empirical results showing a reduction of dimensionality of neural activity with increased stimuli [31]. The situation can be quite different in the range of small to medium input rates, where the mean activity reflects the underlying network organization.

Figure 7 plots average activity and dimensionality in our sample of synthetic networks, as a function of $\langle K \rangle$, for a small input rate. Disconnected networks display similar levels of activity until the percolation threshold is crossed (see inset of Figure 7b). Here, isolated network components display maximum levels of mean activity (because of their lack of dispersal). For the same input load $\lambda = 10^{-4}$, GNC networks are significantly less active (in average) than random networks. That is, low average activity correlates with high dimensionality (Figure 7b).

Efficient transmission in GNC networks appears to be related to dominant negative spins (see Figure 4a). On the other hand, the high activity of random graphs is associated with an abundance of neutral spins and bi-directional links. As shown in the inset of Figure 4b, structural diversity (as measured by spin variance σ_s^2) decays in random networks with increasing average degree. The probability p of linking a pair of nodes in the random graph is constant and independent of the identity of nodes. This is not the case in the GNC model, in which node ordering induces a large spin variance.

4. Discussion

Using the definition of spin and dynamical simulations, our study suggests how global dynamics is constrained by local structures. Structural diversity restraints system' dynamics to a space having a number of dimensions smaller than the maximum degrees of freedom, i.e., network size. Spin heterogeneity (as measured by diversity and structural entropy) can determine the dimensionality of flow networks, that is, the regions of the (D, H) -space are filled according to different structural constraints. These results are consistent with recent studies on the role of local features (motifs) in the dynamical response of spiking neurons. Here, we have shown a similar conjecture by looking at the lowest structural level, the nodes. This is a necessary step towards building a multi-level theory of flow networks, from the smallest level (nodes), to the mesoscale (sub-graphs), to the global scale (full network).

In flow networks, the relationship between in-degree and out-degree limits the computational capacity of nodes. Real networks display correlations in their in-degree and out-degree sequences. This has been associated to the presence of trade-offs, e. g., a tension between complexity and flexibility [33]. For example, in a software system we can find code blocks (also known as "subroutines") with a high number K_1 of incoming connections associated to reusable functions [17]. A large number K_2 of outgoing connections is less frequent in software, possibly because an increased cost of maintenance. Our experiments with GNC networks provide a new dynamical perspective on the tinkered evolution of these artificial systems [34].

Directionality is associated with the persistence of dynamical states and the robustness of the network. Our framework suggests the average lifetime of information (studied in the mean-field approach) relates to large-scale network properties (see Figure 7(a)). A lack of directionality facilitates the emergence of feedback loops and strong correlations. In a random graph, flow is persistently gathered by the GSCC component (where many intermediate nodes are activated) [35]. The situation is reversed for acyclic GNC networks, which continuously push flow towards their sink nodes (i.e., nodes with average spin $\langle s \rangle = 1$). Directionality keeps many nodes underutilized (low network activity), which allows the system to explore a large space of dynamical states (a high number of dimensions).

Future work should determine how directionality affects dynamics using more realistic models, including the impact of structural alterations [36]. Empirical studies indicate the flow of information takes some preferential directions [37], either between individuals (social networks), enzymes (metabolism), species (ecosystems) and brain regions (connectome). The study of these systems defines an idiosyncratic collection of different techniques. Given the interest in understanding how directionality impacts these systems, our methodology provides an unified approach for detecting (any) common mechanisms. This is crucial when predicting the consequences of structural changes, like removal of links or the addition of bidirectional connections, and how directed networks define what states can - or cannot - be reached.

Author Contributions: Conceptualization, S.V.; Methodology, S.V.; Original Draft Preparation, B.V., S.D., S.V., ; Writing-Review and Editing, B.V, S.D.,S.V.; Visualization, B.V., S.V.; Supervision, S.V.

Funding: This study was supported by the Spanish Ministry of Economy and Competitiveness, grant FIS2016-77447-R MINECO/AEI/FEDER and the European Union (SD, SV). SV thanks the CSIC grant "Proyecto Intramural Especial (PIE) 201830/108" (2018-2020) for providing the graphics hardware and virtual reality equipment needed to develop 3-D renderings in this project. During this study, BV was funded by PR01018-EC-H2020-FET-Open MADONNA project.

Conflicts of Interest: The authors declare no conflict of interest.

Data Availability Statement: Results of our analyses can be freely accessed from the website https://github.com/svalver/Structural_Entropy. Network simulation codes will be provided by the authors upon reasonable request.

Abbreviations

The following abbreviations are used in this manuscript:

GNC	Growth Network by Copying model
GWCC	Giant Weak Connected Component
GSCC	Giant Strong Connected Component
GOUT	Giant Out-Component

Acknowledgments: We thank Prof. Joaquin Goñi for inspiring us to investigate the relationship between structure and dynamics. The authors thanks their families and all their collaborators.

References

1. Newman, M. Networks: An Introduction. 2010: Oxford University Press. *Artif. Life* **2012**, *18*, 241–242.

2. Tero, A.; et al.. Rules for Biologically Inspired Adaptive Network Design. *Science* **2010**, 327.
3. Blonder, B.; Dornhaus, A. Time-ordered networks reveal limitations to information flow in ant colonies. *PloS one* **2011**, 6, e20298.
4. Theraulaz, G.; Bonabeau, E.; Deneubourg, J.L. Self-Organization of Hierarchies in Animal Societies. *J. Theor. Biol.* **174**, 313–323.
5. Hilbert, M.; López, P. The world's technological capacity to store, communicate, and compute information. *science* **2011**, 332, 60–65.
6. Valverde, S. Major transitions in information technology. *Philosophical Transactions of the Royal Society B: Biological Sciences* **2016**, 371, 20150450.
7. Valverde, S.; Solé, R.V. Internet's critical path horizon. *Eur. Phys. J. B* **2004**, 38, 245–252.
8. Valverde, S. Breakdown of modularity in complex networks. *Frontiers in physiology* **2017**, 8, 497.
9. Dehmer, M.; Chen, Z.; Emmert-Streib, F.; Tripathi, S.; Mowshowitz, A.; Levitchi, A.; Feng, L.; Shi, Y.; Tao, J. Measuring the complexity of directed graphs: A polynomial-based approach. *Plos one* **2019**, 14, e0223745.
10. Ma'ayan, A.; Cecchi, G.A.; Wagner, J.; Rao, A.R.; Iyengar, R.; Stolovitzky, G. Ordered cyclic motifs contribute to dynamic stability in biological and engineered networks. *Proceedings of the National Academy of Sciences* **2008**, 105, 19235–19240.
11. Debashis, D. *Basic Electronics*; Dorling Kindersley; p. 557.
12. RENYI, E. On Random Graph. *Publicationes Mathematicae* **1959**, 6, 290–297.
13. Erdős, P.; Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **1960**, 5, 17–60.
14. Karrer, B.; Newman, M.E.J. Random graph models for directed acyclic networks. *Phys. Rev. E* **2009**, 80, 046110.
15. Krapivsky, P.L.; Redner, S. Network growth by copying. *Phys. Rev. E* **2005**, 71, 036118.
16. Valverde, S.; Solé, R.V.; Bedau, M.A.; Packard, N. Topology and evolution of technology innovation networks. *Phys. Rev. E* **2007**, 76, 056118.
17. Valverde, S.; Solé, R.V. Logarithmic growth dynamics in software networks. *Europhys. Lett.* **2005**, 72, 858–864.
18. Valverde, S.; Solé, R.V. Punctuated equilibrium in the large-scale evolution of programming languages. *R. Soc. Interface* **2014**, 12, 20150249.
19. Callaway, D.S.; Hopcroft, J.E.; Kleinberg, J.M.; Newman, M.E.J.; Strogatz, S.H. Are randomly grown graphs really random? *Phys. Rev. E*, **64**, 041902.
20. Ahajjam, S.; Badir, H. Identification of influential spreaders in complex networks using HybridRank algorithm. *Scientific reports* **2018**, 8, 1–10.
21. Cimini, G.; Squartini, T.; Garlaschelli, D.; Gabrielli, A. Systemic risk analysis on reconstructed economic and financial networks. *Scientific reports* **2015**, 5, 1–12.
22. Dorogovtsev, S.N.; Mendes, J.F.F.; Samukhin, A.N. Giant strongly connected component of directed networks. *Physical Review E* **2001**, 64, 025101.
23. Boguñá, M.; Serrano, M.A. Generalized percolation in random directed networks. *Phys. Rev. E*, **72**, 016106.
24. Snijders, T.A. The degree variance: an index of graph heterogeneity. *Social networks* **1981**, 3, 163–174.
25. Smith, K.M.; Escudero, J. Normalised degree variance. *Applied Network Science* **2020**, 5, 1–22.
26. Wilson, H.R.; Cowan, J.D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical journal* **1972**, 12, 1–24.
27. Carlu, M.; Chehab, O.; Dalla Porta, L.; Depannemaecker, D.; Héricé, C.; Jedynak, M.; Ersöz, E.K.; Muratore, P.; Souihel, S.; Capone, C.; others. 50 Years of Modeling Neural Activity: Celebrating Jack Cowan's Career: A mean-field approach to the dynamics of networks of complex neurons, from nonlinear Integrate-and-Fire to Hodgkin-Huxley models. *Journal of Neurophysiology* **2020**, 123, 1042.
28. Piantadosi, S.T. One parameter is always enough. *AIP Advances* **2018**, 8, 095118, [<https://doi.org/10.1063/1.5031956>]. doi:10.1063/1.5031956.
29. Cayco-Gajic, N.A.; Clopath, C.; Silver, R.A. Sparse synaptic connectivity is required for decorrelation and pattern separation in feedforward networks. *Nature Communications* **2017**, 8, 1116.
30. Litwin-Kumar, A.; Harris, K.D.; Axel, R.; Sompolinsky, H.; Abbott, L.F. Optimal Degrees of Synaptic Connectivity. *Neuron* **2017**, 93, 1153–1164.e7.
31. Mazzucato, L.; Fontanini, A.; La Camera, G. Stimuli reduce the dimensionality of cortical activity. *Frontiers in systems neuroscience* **2016**, 10, 11.
32. Recanatesi, S.; Ocker, G.K.; Buice, M.A.; Shea-Brown, E. Dimensionality in recurrent spiking networks: global trends in activity and local origins in connectivity. *PLoS computational biology* **2019**, 15, e1006446.
33. Maslov, S.; Sneppen, K. Specificity and stability in topology of protein networks. *Science* **2002**, 296, 910–913.
34. Solé, R.; Valverde, S. Evolving complexity: how tinkering shapes cells, software and ecological networks. *Philosophical Transactions of the Royal Society B* **2020**, 375, 20190325.
35. Wright, E.A.; Yoon, S.; Ferreira, A.L.; Mendes, J.F.; Goltsev, A.V. The central role of peripheral nodes in directed network dynamics. *Scientific reports* **2019**, 9, 1–11.
36. Curado, M.; Escolano, F.; Lozano, M.A.; Hancock, E.R. Early Detection of Alzheimer's Disease: Detecting Asymmetries with a Return Random Walk Link Predictor. *Entropy* **2020**, 22, 465.
37. MacKay, R.; Johnson, S.; Sansom, B. How directed is a directed network? *Royal Society open science* **2020**, 7, 201138.