Relating the small world coefficient to the entropy of 2D networks and applications in neuromorphic engineering

To cite this article: V Onesto *et al* 2019 *J. Phys. Commun.* **3** 095011

View the article online for updates and enhancements.

Journal of Physics Communications

PAPER

Relating the small world coefficient to the entropy of 2D networks and applications in neuromorphic engineering

V Onesto[1] , M Romano[2], F Gentile[3,4] and F Amato[3,4]

[1]   Center for Advanced Biomaterials for Health Care, Italian Institute of Technology, 80125 Naples, Italy
[2]   Department of Experimental and Clinical Medicine, University Magna Graecia, 88100 Catanzaro, Italy
[3]   Department of Electrical Engineering and Information Technology, University Federico II, 80125 Naples, Italy
[4]   These authors share senior authorship.

E-mail: francesco.gentile2@unina.it

Abstract

The study of networks pervades science. The techniques of networks are recently being applied to biomedical disciplines, where the complexity of biomedical systems requires new schemes that can process elevated volumes of data with high efficiency. Artificial neural networks, on which much of artificial intelligence relies, are statistical models partially modeled on biological neural networks. They are capable of modeling and processing nonlinear relationships between inputs and outputs in parallel—in opposition to deterministic models and classical computation schemes, which perform tasks in linear sequences of calculations and may fail to keep up with the challenges of complex biological systems. For these biological or bio-inspired systems, the performance of the networks depends on their topological characteristics. Here, we generated a large number of configurations of points in a plane, in which the entropy s of the configurations was varied over large intervals. Then, we connected points using the Waxman model to obtain the corresponding networks. In correlating the entropy (s) to the small-world coefficient (SW) of those networks, we found that SW varies hyperbolically with s as $\mathrm{SW} = 0.88 + 0.28/s$, where s is expressed in millibits per node. Since the entropy of a distribution of points depends in turn on the density of those points in the plane, such a relationship suggests that the distribution of mass (s) in a complex system determines the topological characteristics (SW) of that system. The small-world-ness of the system, in turn, determines its information efficiency. These findings may have implications in neuromorphic engineering, where chips modeled on biological brains may lead to machines that are able, as for some examples, to diagnose diseases, develop drugs and drug delivery systems faster, design personalized treatments targeted to patient's needs.

1. Introduction

Small world networks are networks with high values of clustering coefficient of the nodes of the networks and very short paths between them (Watts and Strogatz 1998, Strogatz 2001, Watts 2003). Small world networks typically feature over-abundance of hubs with a high number of connections: recently, they have sparked interest because it is believed that networks with a small world topology may feature enhanced signal propagation speed and computational capabilities compared to regular, periodic or random grids of the same size. Small world graphs lie between the extremes of order and randomness (Watts 2003, Crutchfield 2012)—they are used to model dynamical systems where the efficiency of the systems depend less on the components of the system taken in isolation and more on the fact that a large number of components form complex networks. Examples of small world networks have been reported in fields such as cell biology and neuroscience (Achard *et al* 2006, Bullmore and Sporns 2012), theoretical virology (Moore and Newman 2000), systems dynamics (Lago-Fernández *et al* 2000), in the analysis of the topological characteristics of the world wide web and social

Figure 1. Starting from uniform distributions, points having an increasing degree of clustering were generated by varying ξ between 0.2 and 5.
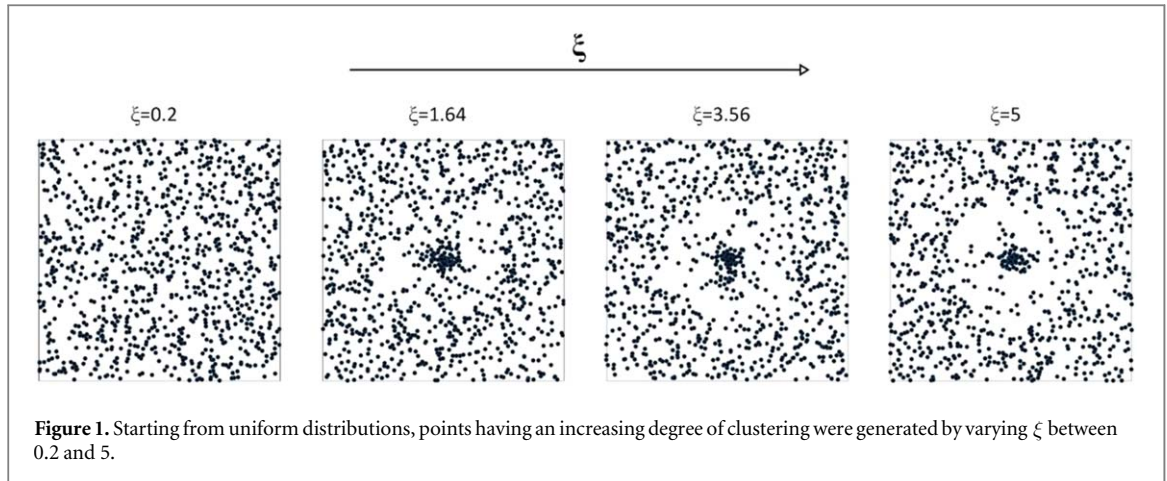
networks (Comellas *et al* 2000). In other reported studies (Takahashi *et al* 2010), it has been demonstrated that the functional and anatomical connectivity among individual neurons exhibits small-world architectures. In references (Marinaro *et al* 2015, Onesto *et al* 2017), some of the authors of the present paper used surfaces with controlled nanotopography to guide the organization of neuronal cells into small world networks with enhanced information flows. Using experiments, information theory approaches and network analysis, we have demonstrated that the formation of the fundamental computation units of the nervous system (such as cortical mini-columns in the cerebral cortex) is guided by the interplay between energy minimization, information optimization and topology. For these systems, the biological functions are determined by the geometrical form, structure and size of the systems themselves.

Here, we correlated the entropy s of distributions of nodes in a plane to the small world coefficient SW of the corresponding networks obtained by wiring those nodes with a probabilistic Waxman model. Moreover, using information theory variables and computer simulations, we found the amount of information transported in a grid as a function of s. These $s - SW$ maps can be used as a preliminary mathematical reference to determine the topological characteristic of a structure without direct knowledge of its internal connections. This study may have implications in bio computing, biosensors operations and neural cell based sensors, the diagnosis and analysis of neurodegenerative disorders, neural development (Decuzzi and Ferrari, 2010, Chiappini *et al* 2015, Onesto *et al* 2017), the analysis of cell-surface interactions and problems at the bio interface, the assembly of cells into complex structures.

2. Experimental section

2.1. Generating clustered distributions of points in the plane

We generated sets of points in the plane with different values of density. To do this, we placed in a bounded domain 1000 points where the coordinates of the points were randomly picked from a uniform distribution. Then, we chose in the domain a fixed position (center) toward which the entire system was left free to evolve. Then, for each point (x_o, y_o) of the initial set, we recalculated its coordinates as

$$
\begin{aligned}
x_f &= c_x + \delta \cos \vartheta \\
y_f &= c_y + \delta \sin \vartheta
\end{aligned}
\tag{1}
$$

where (c_x, c_y) are the coordinates of the center, $\delta = H(\delta^o/l)^\xi$, δ^o is the initial distance between the considered point and the center, l is a cut off distance chosen as 0.25 times the length of the domain, $H = 1$ is a constant (Gentile *et al* 2013). Thus points are displaced proportionally to their distance to the center and to the exponent ξ. ξ was varied in the $0.2 - 5$ intervals to obtain distributions with an increasing degree of clustering (figure 1). By varying the number of centers between 1 and 4, we generated different series of points in the plane (figure 2).

2.2. Determining the spatial density of node distributions

In order to extract the entropy of the distributions, for each node configuration we firstly derived the corresponding density.

To do this, we counted the number n of points falling within a cut-off distance δ_c from a specified point i and divided the result by the sum of the n_i's computed over all the nodes of the set N: $p(i) = n_i / \sum_i^N n_i$. Thus $p(i)$ is the density of points in the neighborhood of i (figure 3). We repeated the process iteratively for each point in the distribution. Notice that, by construction, p is valued between 0 and 1, moreover, the values of $p's$ sum up to
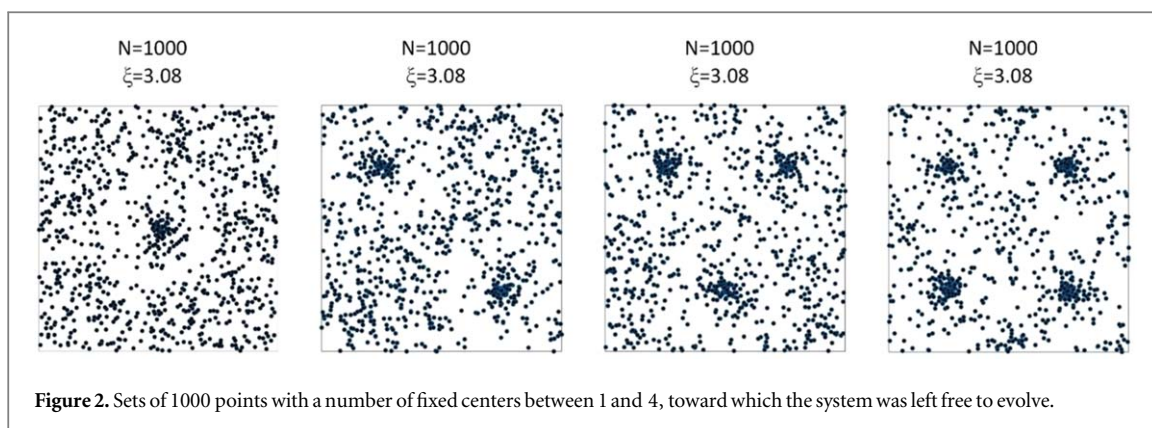
Figure 2. Sets of 1000 points with a number of fixed centers between 1 and 4, toward which the system was left free to evolve.
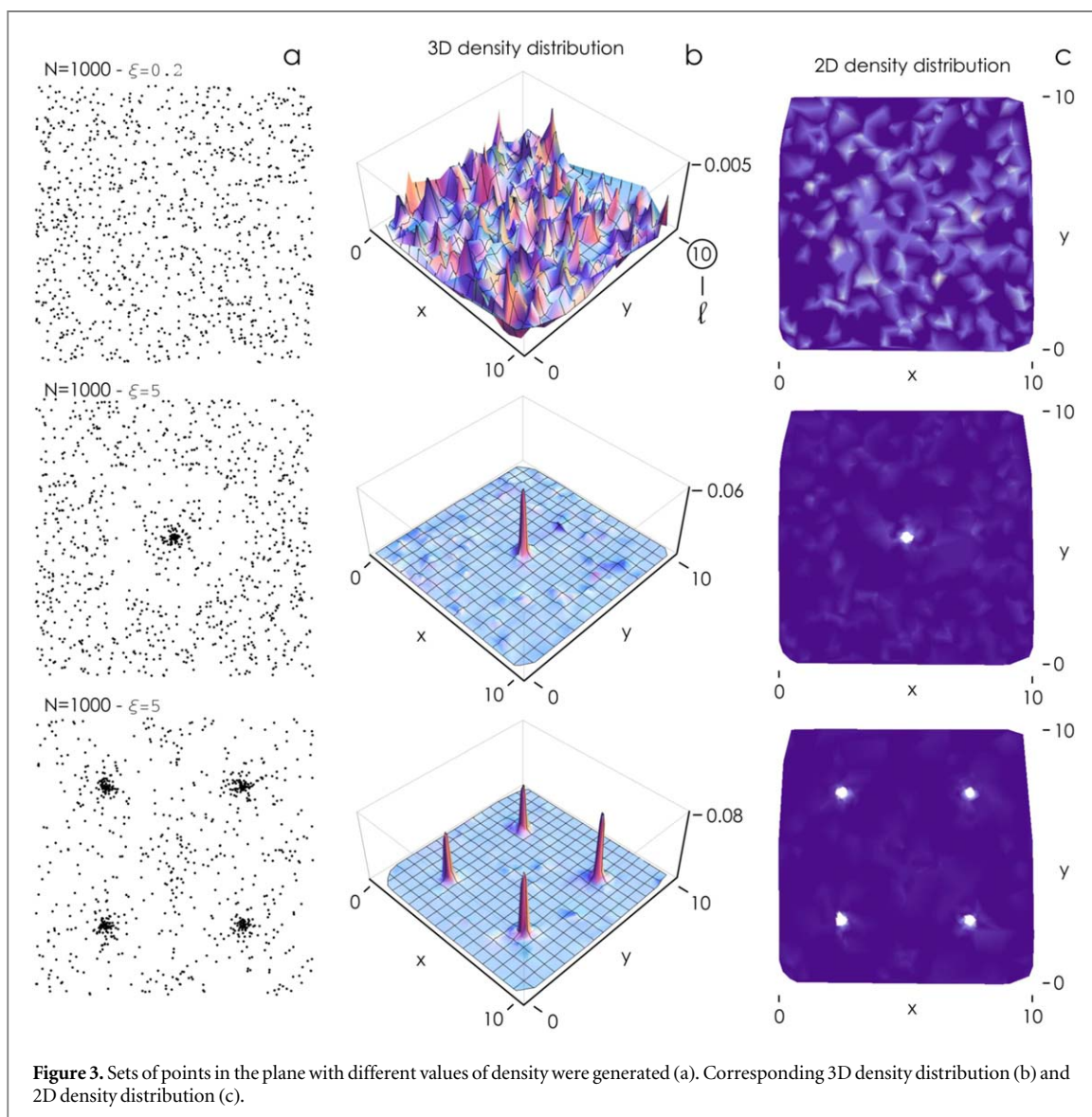

Figure 3. Sets of points in the plane with different values of density were generated (a). Corresponding 3D density distribution (b) and 2D density distribution (c).

unity: therefore $p(i)$ is the probability of finding a node in the neighborhood of i. For the present configuration, we set δ_c as $\delta_c = w/\sqrt{N}$, where w is the initial length of the domain. This value for δ_c enables large $n's$ and statistical significance of the analysis, still ensuring that the $p(i)$ is evaluated finely without loss of information on the local scale.

2.3. 2.3. *Determining the entropy of node distributions*

Upon evaluation of the probability function $p(i) = p_i$, we determine the entropy associated to specific distributions of nodes using the definition of entropy given by Shannon(Shannon 1948)

$$S(set) = -\sum_i p_i \log_2 p_i \tag{2}$$

i.e., the measure of entropy associated with each possible data value is the negative logarithm of the probability mass function for the value. Notice that entropy is measured in bits if the logarithm is taken with base 2. From equation (2) it readily follows that, for a same initial number of points N, intermediate densities have higher entropy than distributions with lower and higher density values. Thus, clustered sets of points with high $p's$ have —typically –lower values of entropy than uniform distributions of points with small $p's$.

Generally, entropy refers to disorder or uncertainty, and the definition of entropy used here (Shannon information entropy) is directly analogous to the definition used in statistical thermodynamics and the Boltzmann's (Campisi and Kobe 2010):

$$S(state) = K_b \ln(\Omega) \tag{3}$$

and Gibbs's (Swendsen 2008, Županovi and Kui 2018)

$$S(set) = -K_b \sum_i p_i \ln p_i \tag{4}$$

equations, where K_b is the constant of Boltzmann, and Ω is the number of microstates (various combinations of particles in various energy states) that can yield the given macrostate.

2.4. Connecting nodes through the Waxman algorithm

From the distributions of points in the plane we determined the corresponding graphs by connecting nodes by a wiring algorithm. We used the Waxman model (Waxman, 1988, Marinaro *et al* 2015), whereby the probability of being a link between two nodes exponentially decreases with the Euclidean distance between those nodes. For a given set of two nodes u and v, the link probability, $P(u, v)$ is defined as:

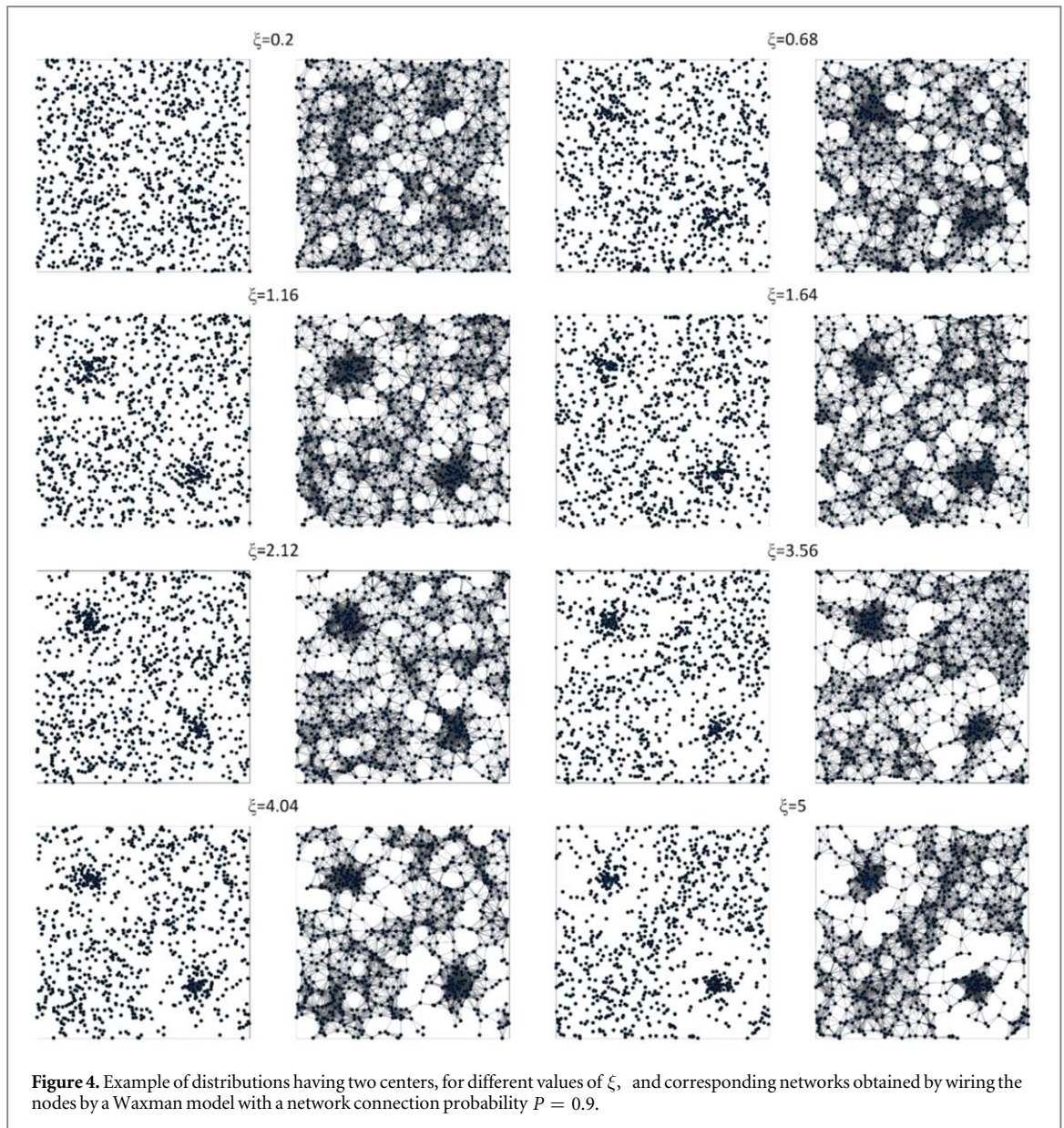$$P(u, v) = \alpha e^{-d(u,v)/\beta L} \tag{5}$$

where d is the Euclidean distance between nodes u and v, and L is the largest possible Euclidean distance between two nodes of the grid. In the equation, α and β are the Waxman model parameters and, upon tuning these, the graph may be more or less dense. α and β should be chosen between 0 and 1. Selecting smaller values of these parameters results in a smaller number of links. For the present configuration, these parameters were set to $\alpha = 1$ and $\beta = 0.025$. The probability P varies between 0 for a pair of nodes with an ideally infinite distance, and 1 for a pair of nodes with an ideally zero distance. The information about the connections among the nodes in a graph is contained in the *adjacency matrix* $A = a_{ij}$, where the indices i and j run through the number of nodes N in the graph; $a_{ij} = 1$, if there exists a connection between i and j, $a_{ij} = 0$ otherwise (Chartrand and Zhang, 2012, Barabási, 2016). In the analysis, reciprocity between nodes is assumed, and thus if information can flow from i to j, it can reversely flow from j to i. In the framework of graph theory, we call a similar network an undirected graph. Notice that this property translates into symmetry of A being $a_{ij} = a_{ji}$. Moreover, $a_{ii} = 0$. We showed above how to derive the distances between nodes d_{ij} in the networks. On the basis of d, we may decide whether a pair of nodes is connected, we use at this end the formula:

$$\alpha e^{-d_{i,j}/\beta L} - R \geqslant 0 \tag{6}$$

in which R is a constant that we have chosen being 0.05 and 0.2 so that the probability of being a connection varies, for different configurations, between $P = 0.95$ and $P = 0.8$.

Equation (5) describes the probability P of being a link between two nodes; as such, it is comprised between 0 and 1: it is the likelihood that two neurons establish a connection in real, biological networks, based on their distance. Then, by comparing these values of probability to a threshold R, arbitrarily chosen between 0 and 1, we make a decision on whether nodes of the network are connected ($P > R$) or not ($P < R$). The comparison, operated through equation (6), makes a probability collapse into a deterministic value, similarly to the wave function collapse in quantum mechanics. The smaller R, the higher the number of connections between nodes of the networks. The Waxman model makes an *hypothesis* on the probability of connection between neurons as a function of their distance. In reference (Ercsey-Ravasz *et al* 2013), from experimental data of cortical connectivity in the macaque, it has been deduced a law that describes the probability f_{ij} of a neuronal projection between two neurons, i and j, as $f_{ij} = c \, e^{-\lambda \, d_{ij}}$, where c and λ are some constant values, and d_{ij} is the distance between them. The form of this equation being justified by the fact that projections at longer distances come at a metabolic cost for individual neurons.

Remarkably, the Waxman model that we used in the present study, *incorporates* the data analyzed by Ercsey-Ravasz and colleagues in their 2013 article. The form of the dependence between the (i) node to node distance

Figure 4. Example of distributions having two centers, for different values of ξ, and corresponding networks obtained by wiring the nodes by a Waxman model with a network connection probability $P = 0.9$.

and the (ii) inter-nodal connection is the same in the model as in the experimental data provided by Ercsey-Ravasz and colleagues, being an exponential decay. However, differently from the experimental data—that are relative to very special configurations or conditions, i.e. cortical connectivity in the macaque—the parameters of the Waxman model enables the model to reproduce a variety of different systems, being more general in scope than a simple set of data.

Figure 4 shows an example of distributions of points in the plane routed by the Waxman algorithm. Points were generated by gradually shifting points, sampled from initially uniform distributions, towards two accumulation points (cluster centers) to an extent proportional to the exponent ξ. The parameter ξ was varied between 0.2 and 5. Points were then connected using a probability of connectivity $P = 0.9$.

2.5. Network analysis

For the generated graphs, we quantified their network parameters, i.e. the clustering coefficient (C_c), the characteristic path length (cpl) and the small world coefficient (SW).

In graph theory, the clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. C_c ranges from 0 (none of the possible connections among the nodes are realized) to 1 (all possible connections are realized and nodes group together to form a single aggregate). The clustering coefficient is defined as (Chartrand and Zhang 2012, Barabási 2016):

$$C_i = \frac{2E_i}{k(k-1)} \qquad (7)$$

where k is the number of neighbors of a generic node i, E_i is the number of existing connections between those, $k(k-1)/2$ being the maximum number of connections, or combinations, that can exist among k nodes. Notice that the clustering coefficient C_i is defined locally - the *global* value C_c is derived upon averaging C_i over all the nodes that compose the graph. The characteristic path length is defined as the average number of steps along the shortest paths for all possible pairs of network nodes (Chartrand and Zhang, 2012, Barabási, 2016). We shall call the minimum distance between a generic couple of nodes the *shortest path length* (Spl), which is expressed as an integer number of steps. Here, we calculate the Spl between any combination of nodes n_l and n_m using the Seidel's algorithm (Seidel, 1995). The algorithm accepts as input the *adjacency matrix A* and produce as output a matrix D where the elements of D, D_{ij}, represent the length of the shortest path from vertex i to vertex j in the graph. Then, the *characteristic path length Cpl* is calculated like the average of Spl over D. The algorithm is a classical, 1992 solution for the All-Pairs-Shortest-Path (APSP) problem for unweighted undirected graphs: it finds path-lengths recursively by the power of the adjacency matrix. It is based on the observation that the product $a_{ik}a_{kj}$ is 1 if there is a path of length 2 from j to i via k, and 0 otherwise. The total number of paths of length 2 from j to i, via any other vertex, is $\sum_k^N a_{ik}a_{kj} = [A \cdot A]_{ij} = A_{ij}^2$. Generalizing to paths of arbitrary length r, we find that A_{ij}^r is the number of paths of length r that connect j to i. The algorithm necessitates to multiply the adjacency matrix by itself repeatedly: it solves the APSP problem in a time $O(M(N)\log(N))$, where $M(N)$ denotes the time necessary to multiply two $N \times N$ matrices of small integers, that in turn is known to be $o(N^{2.376})$.

Once obtained the C_c and Cpl values, we can define a precise measure of small-world-ness, the small world coefficient (SW), based on the trade off between high local clustering and short path length (Humphries and Gurney 2008, Narula *et al* 2017).

A network G with n nodes and m edges is a small-world network if it has a similar path length but greater clustering of nodes than an equivalent Erdos-Rényi (E–R) random graph with the same m and n (an E–R graph is constructed by uniquely assigning each edge to a node pair with uniform probability) (Watts and Strogatz 1998, Strogatz 2001, Watts 2003). Let Cpl_u and Cc_u be the mean shortest path length and the mean clustering coefficient for the E–R random graphs, obtained averaging the Cpl and the Cc of 20 uniform distributions, and Cpl_{graph} and Cc_{graph} the corresponding quantities for the graphs derived using the methods described above. We can calculate:

$$\gamma = \frac{Cc_{graph}}{Cc_u} \qquad (8)$$

$$\lambda = \frac{Cpl_{graph}}{Cpl_u} \qquad (9)$$

Thus, the small world coefficient is

$$SW = \frac{\gamma}{\lambda} \qquad (10)$$

The categorical definition of small-world network above implies $\lambda \geqslant 1$, $\gamma \gg 1$ which, in turn, gives $SW > 1$.

2.6. Simulating information flows in 2D networks of neuronal units

We used a generalized leaky integrate and fire model (FitzHugh 1955, de la Rocha and Parga 2005) to simulate information flow in bi-dimensional neural networks as described in Reference (Onesto *et al* 2016) and recapitulated in this section. Nodes of a grid (network) are generated following the methods described above and in the rest of the article. Each node in the grid is neuronal unit (computational unit) able to receive, elaborate and transmit a signal to another neuronal unit in the grid. The temporal sequence of spikes that propagate along the grid encodes the information transmitted over the entire network. The typical signal transmitted by a neuronal unit over time is train of spikes (action potentials) that may be interpreted using information theory approaches (Strong *et al* 1998, Borst and Theunissen 1999, Quiroga and Panzeri 2009). We represent the variability of individual neurons in response to a long random stimuli sample with the total entropy H. Similarly, the noise entropy N is the variability of the spike train in response to a sample of repeated stimuli. The information content provided by the different spike trains is the difference between entropies: $I = H - N$. We used a generalized leaky integrate and fire model (FitzHugh, 1955; de la Rocha and Parga, 2005) to simulate trains of signals in bi-dimensional networks as described in Reference (Onesto *et al* 2016). Some nodes were randomly selected from the network and excited with a random and periodic signal of time. Upon excitation, spikes propagate in cascade in the grid. To simulate the flow of signal in the networks, we used a generalized leaky integrate and fire model (FitzHugh, 1955; de la Rocha and Parga, 2005). In individual neurons, electric pulses

excite the neuron until the response (potential) at the postsynaptic sites reaches and surpasses a limiting value (threshold potential), then, the target neuron produces an impulse (an action potential) that propagates in turn to another neuron. This process is described by the following equation, in which the membrane potential V obeys to a function of the sole time:

$$C_m \frac{dV}{dt} = -g_l(V - V_o) + I_{stim} \tag{11}$$

Where C_m is the capacitance of the membrane, g_l is the conductance, V_o is the resting potential of the neuron. The current I_{stim} is the stimulus that excites the neuron until the membrane potential reaches a threshold V_{th} and an action potential is released from the system. Neurons in a grid are described by a set of coupled differential equations that generalizes the model described by equation (11). Each node in the network sends and receives information and this process is mediated through the integrate and fire model and equation (11). Assuming linearity, I_{stim} is given by the superposition of current pulses J generated by the neurons i that fire on a neuron j

$$I_{stim}(t) = \sum_{\mathcal{B}}^{i} \zeta(d_{ij}) J \sum_{rel}^{k} \delta(t - t_i^k) \tag{12}$$

Where rel is the number of neurotransmitter release events, δ is the Dirac delta function, t_i^k is the timing of individual pulses. In equation (12), ζ is a damping term which accounts for the inter-nodal distance d_{ij}. Pulses repeatedly excite a neuron until $V = V_{th}$ and an action potential is discharged from the target neuron. The action potential generates in turn an impulse that propagates through the network. The generation of an action potential at a node of the grid at a specific time is registered as an event. The temporal sequence of events encodes the information transmitted over that grid. Resulting patterns of multiple spike trains are interpreted using information theory approaches (Strong *et al* 1998; Borst and Theunissen, 1999; Quiroga and Panzeri, 2009). Time spikes are grouped in sets of words, in which a word is an array of on (presence of a spike)/off (absence of a spike) events in a binary representation. On sorting words in order of decreasing occurrence in the train, one can derive the associated Shannon entropy H as

$$H(\psi) = -\sum_{\psi} \vartheta(\psi) \log_2 \vartheta(\psi) \tag{13}$$

that quantifies the average amount of information gained with each stimulus presentation. Entropy is measured in bits if the logarithm is taken with base 2. In the equation, $\vartheta(\psi)$ represents the probability with which a stimulus ψ is presented in the set. If H is the variability of individual neurons in response to a long random sample of stimuli (*total entropy*), and N is the variability of the spike train in response to a sample of repeated stimuli (*noise entropy*), then information that the spike train provide about the input is the difference between entropies $I = H - N$. This permits to derive information over all the nodes of the graph.
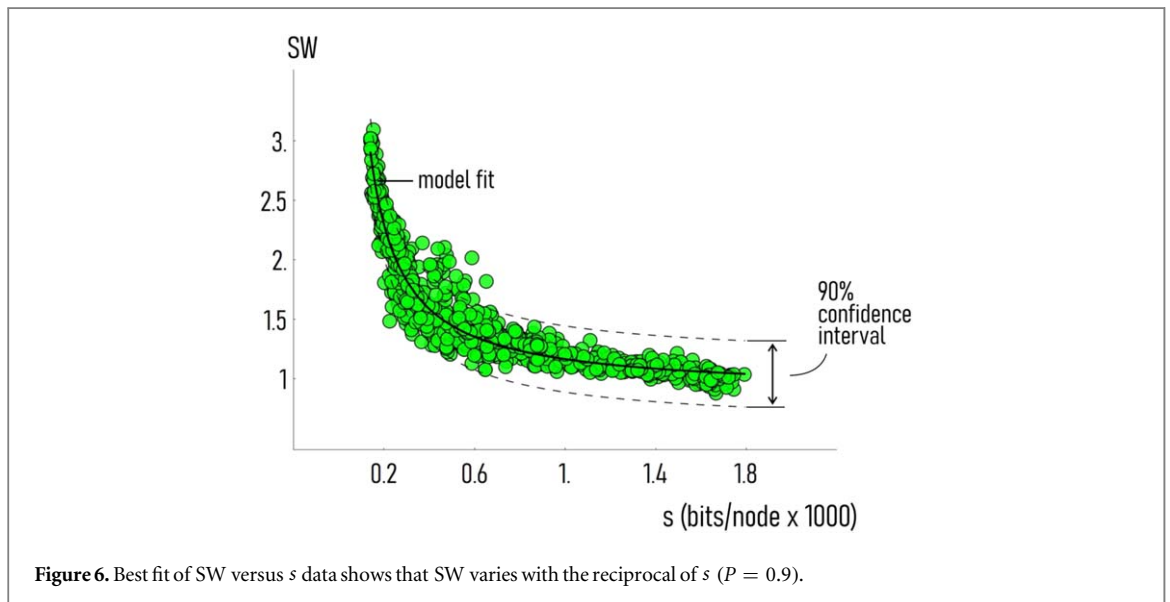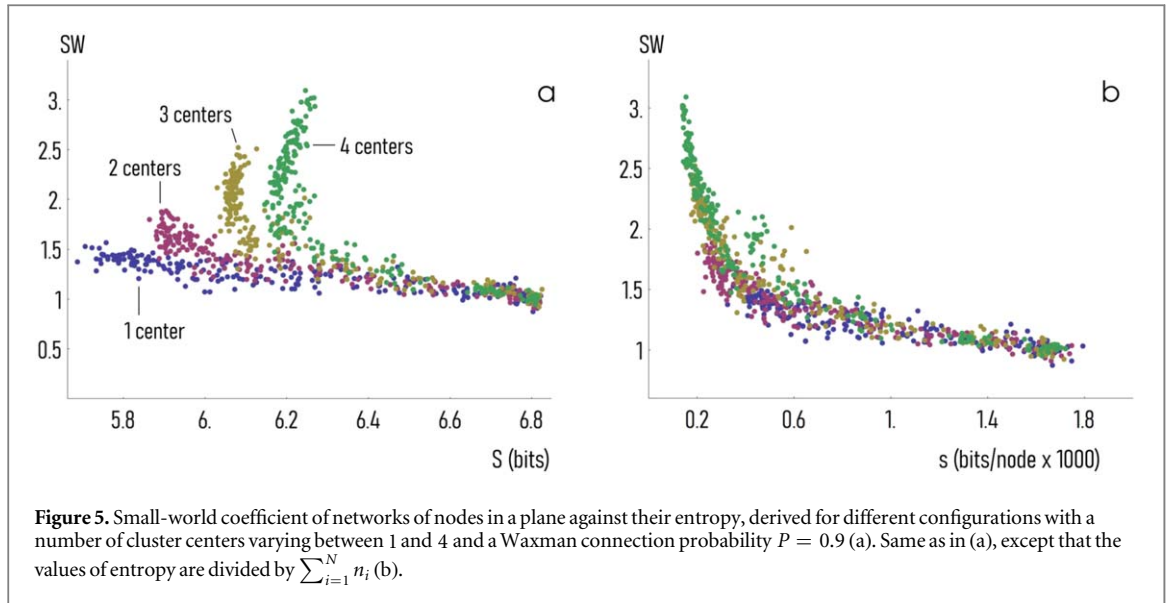
3. Results

3.1. Small world ness of distributions as a function of entropy

We report in figure 5(a) the small-world coefficient SW of networks of nodes in a plane against their entropy S, derived using the methods described above for a large number of different configurations, for the exponent ξ varying between 0.2 and 5 and a Waxman connection probability $P = 0.9$. We observe that, while generally SW decreases with S, the larger the number of centers in a distribution, the higher the values of SW associated to a specific value of entropy S (figure 5(a)). The form of the SW(S) curves depends on the number of clusters for which the sets of points were generated, owing to fact that, for lower values of entropy, the points in the distributions are tightly packed together, that in turn leads to overestimate their density p when those points are distributed around one or very few accumulation points. To generate a relationship where dependence between variables does not depend on the configuration used to determine it, for each configuration we normalized S to $\sum_{i=1}^{N} n_i$: $s = S/\sum_{i=1}^{N} n_i$ (figure 5(b)). In this representation, the scatter-plots of SW against s are the same independently on the configuration used to derive them (figure 5(b)).

In the diagram, the values of s vary between $s \sim 0.2$ *millibits/node*, for which the small-world-coefficient is SW ~ 3 (highly clustered distributions), to $s \sim 1.8$ *millibits/node*, for which SW ~ 1 (uniform distributions). Graphical representation of the small-world-ness against entropy and figure 5(b) indicate that SW varies hyperbolically with s. Best fit of data yields the relationship (figure 6):

$$SW = 0.88 + 0.28/s \tag{14}$$

The form of equation (14) suggests that infinitesimal variations of s are directly transferred to SW. Since the entropy is derived starting from the density ϱ of points in the domain, this also suggests that that the density and the topological properties of the distributions are tightly interwoven - and that the small-world-ness, the clustering coefficient and the characteristic path length of a network are encoded in its mass density function

Figure 5. Small-world coefficient of networks of nodes in a plane against their entropy, derived for different configurations with a number of cluster centers varying between 1 and 4 and a Waxman connection probability $P = 0.9$ (a). Same as in (a), except that the values of entropy are divided by $\sum_{i=1}^{N} n_i$ (b).

Figure 6. Best fit of SW versus *s* data shows that SW varies with the reciprocal of *s* ($P = 0.9$).

evaluated over the entire surface of the network. The first derivative of the function can be readily calculated as $\partial \mathrm{SW}/\partial s = -A/s^2$, with $A = 0.28$ *millibits/node*, that indicates the sensitivity of SW to a change in entropy. In reporting the $\mathrm{SW}(s)$ characteristics in figure 7 for different values of the Waxman probability of connection P, we observe the form of the of the $\mathrm{SW} - s$ relationship is preserved, with SW decreasing with s, while the values of A change, being $A^{P_{0.95}} \sim 0.22$, $A^{P_{0.90}} \sim 0.28$, $A^{P_{0.8}} \sim 0.69$ *millibits/node*. Results suggest that the correlation between small-world-ness and entropy for a given distribution is a function of the number of connections in that distribution, and the relationship between the two may be described by the simple law:

$$SW \times s \propto -3.28\, P \tag{15}$$

The fact that *the small world coefficient* (*SW*) of a set of nodes and the *entropy* (*s*) of those nodes are related, may not be unexpected. The entropy of a system of nodes can be determined from the distribution of the nodes in the domain, that in turn is described by a probability function, i.e. local density. With Waxman, the small world coefficient would depend on the relative position and distance d between nodes, because the probability of two nodes of being connected decays exponentially with d. While these two probability (used in the definition of s and SW) are not *exactly* the same quantity, they both depend on the density of points in the domain. Rather than being a circular argument, this indicates that the density (mass distribution) of a system, described by the sole entropy, influences the connectivity of that system (i.e. the small world coefficient). This paper is an attempt to find a quantitative relationship between these variables. We can comment even further on the form of equation (14). The equation and the graphical representation of the $s - \mathrm{SW}$ relationship in figure 7, present an asymptotic behavior in the limit of small ($s \to 0$) and large ($s \to \infty$) s. In the limit of large entropies ($s \to \infty$),

Figure 7. The diagrams show the small-world coefficient SW correlated to the entropy *s* of distributions of nodes in a plane, for different Waxman connection probabilities.
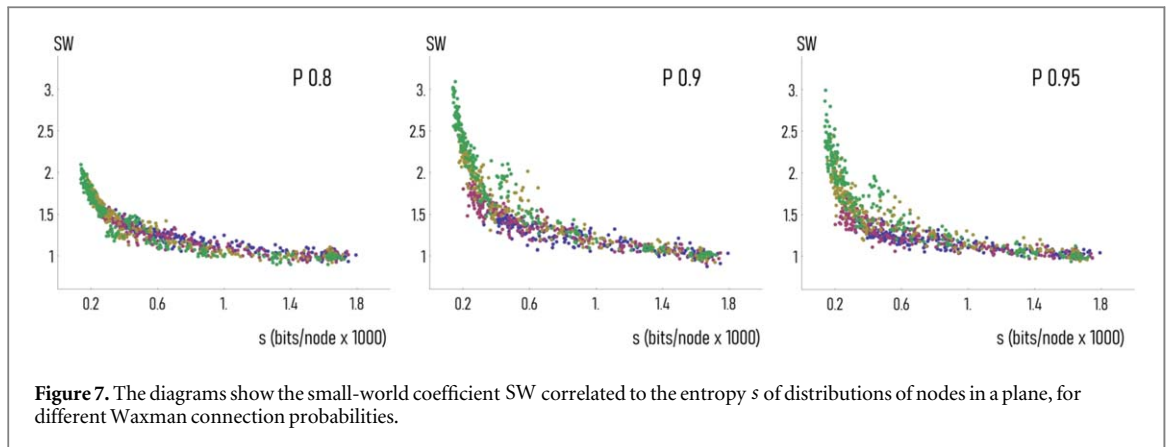
the system tends to a network resulting from a random distribution of points, for which, by mathematical definition of small world coefficient, $SW \equiv 1$. Equation (14) reflects this boundary condition, yielding $SW \sim 1$ for $s \sim 2 \ millibits/node$, and a value of small-world-ness slowly approaching 0.88, $SW \rightarrow 0.88$, for $s \rightarrow \infty$. At the other extreme, equation (14) predicts $SW \rightarrow \infty$ for $s \rightarrow 0$. In the limit of vanishingly small entropies, $(s \rightarrow 0)$, nodes in the domain gather around few accumulation points, with extremely high values of density, resulting in high values of clustering coefficient and small values of characteristic path length, that in turn yields very large values of small-world-ness. Thus, the relationship that we found between *s* and SW is accurate for intermediate values of *s*, and may be less accurate at the extremes of *s*. The accuracy of the formula at low *s* may be increased using an adaptive geometry, i.e. values of cut off distance δ_c (used in the definition of density) that changes as a function of the local density $p(i)$.

3.2. Benchmarking the model

To verify the model capability to predict the small world ness of networks starting from their entropy values, we artificially generated 5 different distributions of nodes in the plane, reported in figure 8. The first 3 distributions are agglomerates of points in the domain, where the number of agglomerates ranges from 6 for the first set, to 7 for the second set, to 9 for the third configuration. The distributions are generated by sampling the *x* and *y* coordinates from Gaussian distributions where the means are the positions of cluster centers, and the variance is 5. In the last two configurations, points are clustered around a circle (4) and a sine-wave (5): sets of points are generated by moving the center of a Gaussian distribution with variance ˜1 along the trajectory of the curves, and sampling the point-coordinates from that distribution. For all configurations the number of points in the domain is $N = 1000$. Using the methods described above, we derived for the distributions the corresponding values of entropy as $s^1 \sim 0.22$, $s^2 \sim 0.26$, $s^3 \sim 0.36$, $s^4 \sim 0.23$, $s^5 \sim 0.70$ bits per 1000 nodes. For all distributions, we then connected the points in the domain using the Waxman algorithms and a probability of connection $P = 0.9$. Using network analysis algorithms we found the associated values of small-world-ness being $SW^1_{exp} \sim 2.15$, $SW^2_{exp} \sim 1.68$, $SW^3_{exp} \sim 1.98$, $SW^4_{exp} \sim 1.45$, $SW^5_{exp} \sim 1.045$ (figure 8). Expected values of small-world-ness predicted from the model and equation (14) are $SW^1_{theory} \sim 2.16$, $SW^2_{theory} \sim 1.96$, $SW^3_{theory} \sim 1.66$, $SW^4_{theory} \sim 2.10$, $SW^5_{theory} \sim 1.28$ with small deviations between the prediction of the model and the true values of small-world-ness. For this test, the residuals are $e^1 = 0.005$, $e^2 = 0.14$, $e^3 = 0.17$, $e^4 = 0.30$, $e^5 = 0.18$, with a values of chi squared $\chi^2 = \sum e_i^2 / \sigma^2 \sim 2.15$.

3.3. Information transported in the networks

We used the methods reported in references(Onesto *et al* 2016, Onesto *et al* 2017) to evaluate the amount of information exchanged by the different grids as a function of the small-world-ness (entropy) of the grids. Information is associated to the probability of an event. When the outcome of a process or the output of a physical, biological, or chemical system has a low-probability value, the information that the response carries about the stimulus is high. In the simulations, we assumed that nodes of the grid are artificial neurons capable to receive a signal and integrating that signal over time. When the cumulative signal reaches a threshold that can be arbitrarily tuned, the neuron in turn generates a signal that propagates in cascade in the grid. A similar model is called a generalized leaky integrate and fire model (FitzHugh, 1955, de la Rocha and Parga, 2005), a more sophisticated evolution of the model in which single computational components are arrayed in systems with a great many of elements - to form functional neuronal networks - has been recently developed by some of the authors of this paper (Onesto *et al* 2016, Onesto *et al* 2017, Onesto *et al* 2018). In a biological interpretation of the scheme, the signal released by the neuron is an action potential, and the sequence and time pattern of action potentials encodes the information transported in the network (Strong *et al* 1998, Borst and Theunissen 1999,
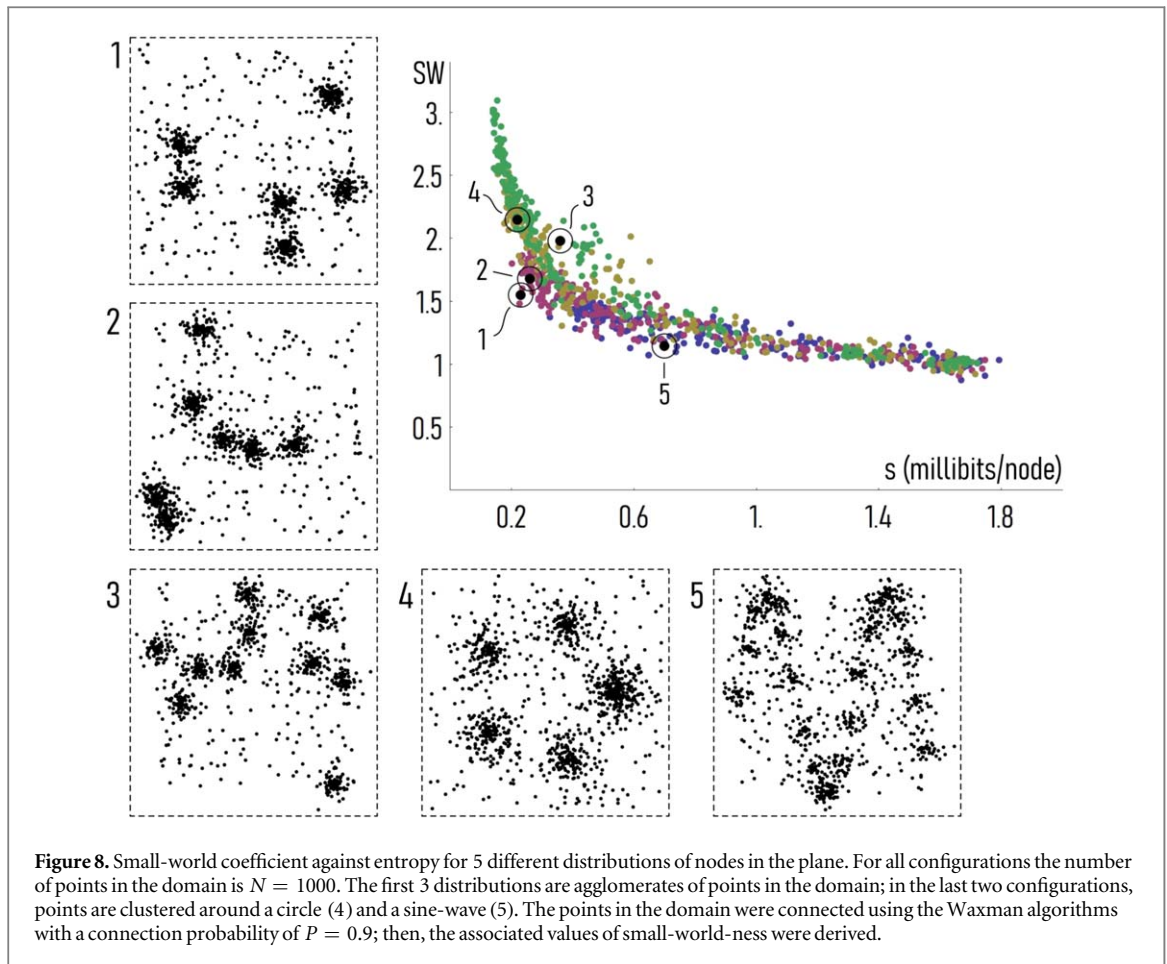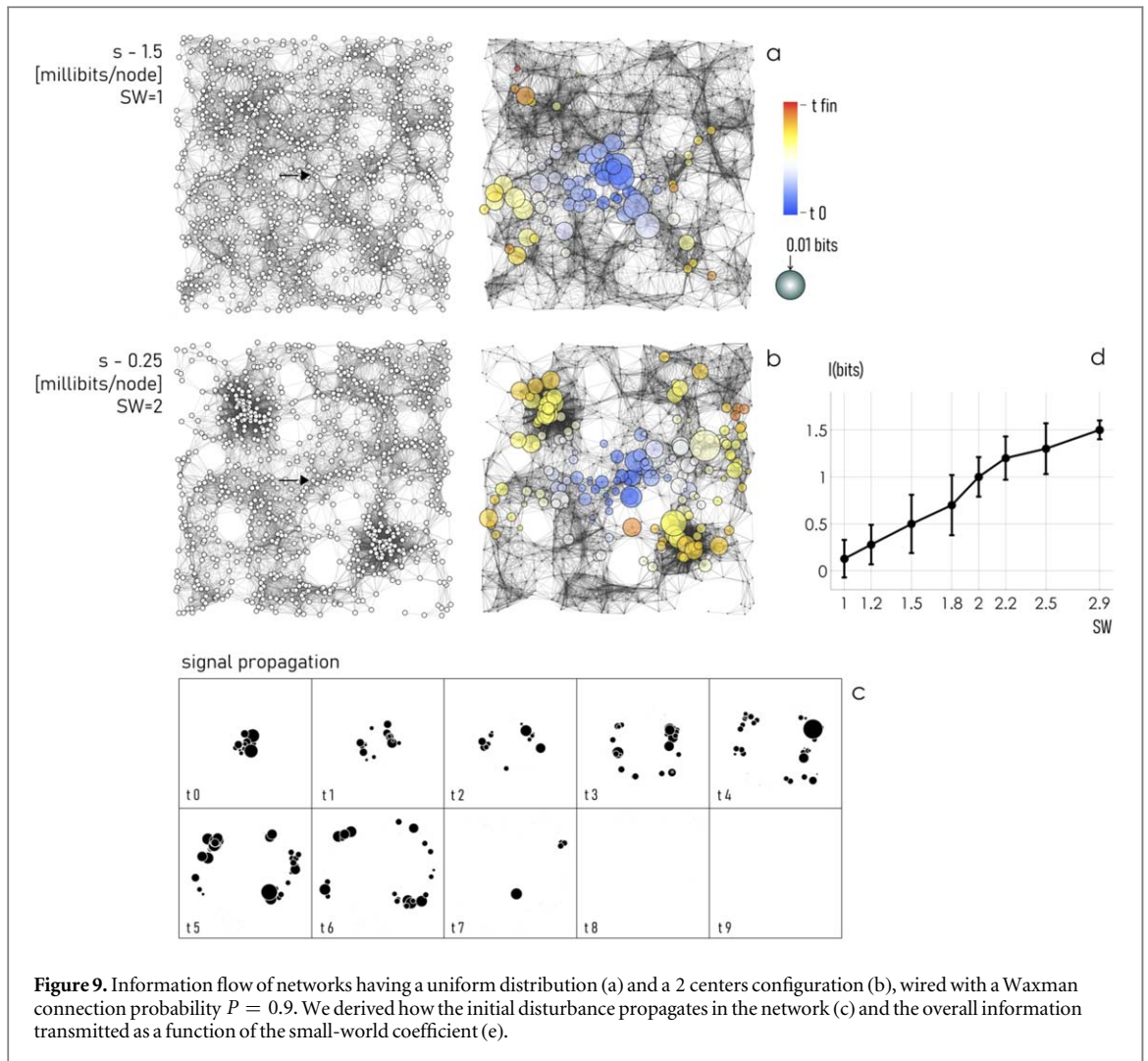
Figure 8. Small-world coefficient against entropy for 5 different distributions of nodes in the plane. For all configurations the number of points in the domain is $N = 1000$. The first 3 distributions are agglomerates of points in the domain; in the last two configurations, points are clustered around a circle (4) and a sine-wave (5). The points in the domain were connected using the Waxman algorithms with a connection probability of $P = 0.9$; then, the associated values of small-world-ness were derived.

Quiroga and Panzeri 2009). To decode information, one can use information theory approaches. The information content of the system can be derived as the difference between the total entropy H and the noise N: $I = H - N$. The total entropy is the variability of individual neurons in response to a long random stimuli sample (Strong *et al* 1998, Borst and Theunissen, 1999, Quiroga and Panzeri, 2009). The noise entropy is the variability of the spike train in response to a sample of repeated stimuli. The entropy (sometimes called Shannon information entropy) of a signal can be determined as $H(\varsigma) = -\sum_{\psi} \vartheta(\psi) \log_2 \vartheta(\psi)$, where $\vartheta(\psi)$ denotes the probability with which a stimulus ψ is presented in the response of the node to the disturbance ς (Strong *et al* 1998, Borst and Theunissen, 1999, Quiroga and Panzeri, 2009;). H quantifies the average amount of information gained with each stimulus presentation. Using this framework, we verified the ability of neuronal networks to elaborate information as a function of their topological properties. We simulated the propagation of a disturbance from the center throughout the entire extension of the networks for different values of the small world coefficient SW. Figure 9(a) reports patterns of information derived for a uniform distribution of nodes with SW ~ 1 and entropy $s \sim 1.5$ millibits/node. Figure 9(b) reports patterns of information derived for a clustered distribution of nodes with SW ~ 2 and entropy $s \sim 0.25$ millibits/node. In this graphical representation, the diameter of colored circumferences around individual nodes is proportional to the information that at a generic time arrives at those nodes. Because of the characteristics of the network, for the second configuration (SW ~ 2) information propagates far away from the point of application of the initial disturbance - and generally the information transported through the nodes is more intense compared to the information transported in the random network with SW ~ 1. In figure 9(c), we report the time evolution of the signal in the network for SW ~ 2. On can observe that the signal travels from the center to the periphery of the network with continuity—the signal itself is persistent and decreases mildly with the distance from the center of the grid. For certain times of propagation, i.e. $t = 4$, the signal is amplified with respect to the initial disturbance. This is relevant, because the topology of the network can result in information confinement and enhancement, similarly in concept to surface plasmons resonance in nano-optics devices (Stockman, 2008; Gentile *et al* 2014). Moreover, we calculated the total information I transmitted in the networks as a function of the networks characteristics (figure 9(d)). I is determined as the information delivered at a node of the grid integrated over the entire grid and over the whole duration of the process of propagation. We launched more than 50 simulations per configuration. We found that I shows a very high sensitivity to the small world

Figure 9. Information flow of networks having a uniform distribution (a) and a 2 centers configuration (b), wired with a Waxman connection probability $P = 0.9$. We derived how the initial disturbance propagates in the network (c) and the overall information transmitted as a function of the small-world coefficient (e).

coefficient and entropy of the networks. I is low for small values of small-world-ness SW ($I^{SW_1} \sim 0.2$ bits), and increases with SW up to $I^{SW_{2.9}} \sim 1.5$ bits in networks with elevated SW $= 2.9$. The maximum enhancement factor of information is $Q_E = I^{SW_{2.9}}/I^{SW_1} \sim 7.5$. Bonferroni post hot test indicates that information transmitted within the networks with SW $= 2.9$ is statistically greater than information transmitted through the networks with SW $= 1$ ($p = 0.05$). Results of this section are relevant in that they show that the distribution of mass of a system may determine its topological characteristics, which, in turn, regulate the information transfer rate and efficiency of that system. Results suggest that from the analysis of the distribution of neurons in a system, one can derive the efficiency of the system, and the maximum information that the system can convey about a stimulus, without direct knowledge of the topology of the networks that they form. With implications in the study of neurodegenerative diseases, regenerative medicine, tissue engineering, neuromorphic engineering. Notice though that here we have not considered techniques to enhance the information transmitted through individual channels, like noise cancellation or compensation, or methods, such as orthogonal frequency division multiplexing, for encoding digital data on multiple carrier frequencies. The introduction of these algorithms, some of which are described in reference (Bibi *et al* 2018), may emphasize even further the importance of *geometry* in networks science and information theory applied to biological systems.

4. Discussion and conclusions

We generated distributions of points in which the density and entropy of the distributions were varied over large intervals. We then used Waxman algorithms to wire nodes and obtain networks with certain degrees of connectivity. In correlating the small world coefficient SW of those networks to the entropy s of the nodes, we found that s and SW are linked by a simple law that, for a Waxman probability of connectivity $P = 0.9$, reads as SW $= 0.88 + 0.28/s$—thus the larger the entropy of the distributions, the smaller the small world coefficient of the corresponding networks. Here SW is expressed in non-dimensional units and s in millibits per node.

Using mathematical models and computer simulations, we verified that networks with high values of small-world-ness and low values of entropy are more information efficient than networks with small values of small-world-ness and high values of entropy. This is relevant because it indicates that the structure and distribution of mass (s) in a complex systems may determine its topological characteristics (SW) that, in turn, influence the information carried by the system. These findings may have implications in neuromorphic engineering, where chips are modeled on biological brains and designed to process sensory data in ways not specifically programmed. Neuromorphic chips attempt to model artificially the massively parallel way the brain processes information—in opposition to the classical von Neumann architecture, which shuttles data between a central processor and memory chips in linear sequences of calculations. In the brain, billions of neurons and trillions of synapses respond to sensory inputs such as visual and auditory stimuli with unmatched efficiency. Those neurons modify their *topology* in response to a change in input—in doing so, they can efficiently process images, sounds and complex inputs that are otherwise untreatable by classical computing methods. Artificial chips that are being developed over time incorporate neural networks to imitate the architecture of the brain. In both, the *topological* characteristics of the neurons activate exceptionally complex functions, such us language, object recognition, intelligence and artificial intelligence, which differentiate - and are beyond the reach - of traditional information schemes. In this scenario, relations like equation (11) that correlate the distribution of the smallest computational elements of a system, to the small world ness and the information flows within that systems, are tools that can be used in the rational design of bio-scaffolds for tissue engineering, regenerative medicine, bio-chips, intelligent medical sensors and devices.

ORCID iDs

V Onesto ⓘ https://orcid.org/0000-0002-2308-4126
F Gentile ⓘ https://orcid.org/0000-0002-1724-6301
F Amato ⓘ https://orcid.org/0000-0002-9053-3139

References

Achard S, Salvador R, Whitcher B, Suckling J and Bullmore E 2006 A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs *the Journal of Neuroscience* **26** 63–72

Barabási A-L 2016 *Network Science* (Glasgow: Cambridge University Press)

Bibi N, Muhammad N and Cheetham B 2018 Inverted wrap-around limiting with bussgang noise cancellation receiver for OFDM signals *Circuits Syst. Signal Process.* **37** 901–14

Borst A and Theunissen F E 1999 Information theory and neural coding *Nat. Neurosci.* **2** 947–57

Bullmore E and Sporns O 2012 The economy of brain network organization *Nat. Rev. Neurosci.* **13** 336–49

Campisi M and Kobe D H 2010 Derivation of the Boltzmann principle *Am. J. Phys.* **78** 608–15

Chartrand G and Zhang P 2012 *A First Course in Graph Theory* (Dover Publications)

Chiappini C, Martinez J O, De Rosa E, Almeida C S, Tasciotti E and Stevens M M 2015 Biodegradable nanoneedles for localized delivery of nanoparticles *in Vivo*: exploring the biointerface *ACS Nano* **9** 5500–9

Comellas F, Ozóna J and Peters J G 2000 Deterministic small-world communication networks *Inf. Process. Lett.* **76** 83–90

Crutchfield J P 2012 Between order and chaos *Nat. Phys.* **8** 17–24

de la Rocha J and Parga N 2005 Short-Term synaptic depression causes a non-monotonic response to correlated stimuli *The Journal of Neuroscience* **25** 8416–31

Decuzzi P and Ferrari M 2010 Modulating cellular adhesion through nanotopography *Biomaterials* **31** 173–9

Ercsey-Ravasz M, Markov N T, Lamy C, Van Esse D C, Knoblauch K, Toroczkai Z and Kennedy H 2013 A predictive network model of cerebral cortical connectivity based on a distance rule *Neuron* **80** 184–97

FitzHugh R 1955 Mathematical models of threshold phenomena in the nerve membrane *Bull. Math. Biol.* **17** 257–78

Gentile F, Coluccio M L, Rondanina E, Santoriello S, Di Mascolo D, Accardo A, Francardi M, De Angelis F, Candeloro P and Di Fabrizio E 2013 Non periodic patterning of super-hydrophobic surfaces for the manipulation of few molecules *Microelectron. Eng.* **111** 272–6

Gentile F, Coluccio M L, Zaccaria R P, Francardi M, Cojoc G, Perozziello G, Raimondo R, Candeloroa P and Fabrizio E D 2014 Selective on site separation and detection of molecules in diluted solutions with superhydrophobic clusters of plasmonic nanoparticles *Nanoscale* **6** 8208–25

Humphries M D and Gurney K 2008 Network 'small-world-ness': a quantitative method for determining canonical network equivalence *PLoS One* **3** e0002051

Lago-Fernández L F, Huerta R, Corbacho F and Sigüenza J A 2000 Fast response and temporal coherent oscillations in small-world networks *Phys. Rev. Lett.* **84** 2578

Marinaro G, La Rocca R, Toma A, Barberio M, Cancedda L, Di Fabrizio E, Decuzzi P and Gentile F 2015 Networks of neuroblastoma cells on porous silicon substrates reveal a small world topology *Integr. Biol.* **7** 184–97

Moore C and Newman M E J 2000 Epidemics and percolation in small-world networks *Phys. Rev. E* **61** 5678

Narula V, Zippo A G, Muscoloni A, Biella G E M and Cannistraci C V 2017 Can local-community-paradigm and epitopological learning enhance our understanding of how local brain connectivity is able to process, learn and memorize chronic pain? *Applied Network Science* **2** 1–28

Onesto V, Cancedda L, Coluccio M, Nanni M, Pesce M, Malara N, Cesarelli M, Fabrizio E D, Amato F and Gentile F 2017 Nano-topography enhances communication in neural cells networks *Scientific Reports* **7** 1–13

Onesto V, Cosentino C, Di Fabrizio E, Cesarelli M, Amato F and Gentile F 2016 Information in a network of neuronal cells: effect of cell density and short-term depression *BioMed Research International* **2016** 1–12

Onesto V, Narducci R, Amato F, Cancedda L and Gentile F 2018 The effect of connectivity on information in neural networks *Integr. Biol.* **10** 121–7

Quiroga R Q and Panzeri S 2009 Extracting information from neuronal populations: information theory and decoding approaches *Nat. Rev. Neurosci.* **10** 173–85

Seidel R 1995 On the all-pairs-shortest-path problem in unweighted undirected graphs *J. Comput. Syst. Sci.* **51** 400–3

Shannon C 1948 A mathematical theory of communication *The Bell System Technical Journal* **27** 379–423

Stockman M I 2008 Spasers explained *Nat. Photonics* **2** 327–9

Strogatz S H 2001 Exploring complex networks *Nature* **410** 268–76

Strong S P, Koberle R, van Steveninck R R d R and Bialek W 1998 Entropy and information in neural spike trains *Phys. Rev. Lett.* **80** 197

Swendsen R H 2008 Gibbs' paradox and the definition of entropy *Entropy* **10** 15–8

Takahashi N, Sasaki T, Matsumoto W, Matsuki N and Ikegaya Y 2010 Circuit topology for synchronizing neurons in spontaneously active networks *Proc. Nat. Acad. Sci. USA* **107** 10244–9

Watts D J 2003 *Small Worlds: The Dynamics of Networks between Order and Randomness* (Woodstock: Princeton University Press)

Watts D J and Strogatz S H 1998 Collective dynamics of 'small-world' networks *Nature* **393** 440–2

Waxman B 1988 Routing of multipoint connections *IEEE J. Sel. Areas Commun.* **6** 1617–22

Županovi P and Kui D 2018 Relation between Boltzmann and Gibbs entropy and example with multinomial distribution *Journal of Physics Communications* **2** 045002