Electronic Supplementary Material (ESI) for Integrative Biology.

The effect of connectivity on information in neural networks

V. Onesto[1,†], R. Narducci[2,†], F. Amato[1], L. Cancedda[2,‡], F. Gentile[3,‡,*]

[1] Department of Experimental and Clinical Medicine, University of Magna Graecia, 88100 Catanzaro, Italy

[2] Istituto Italiano di Tecnologia, Via Morego 30, 16163 Genova, Italy

[3] Department of Electrical Engineering and Information Technology, University Federico II, 80125, Naples, Italy

† These authors contributed equally to this work

‡ These authors shared senior authorship

*author to whom correspondence should be addressed: francesco.gentile2@unina.it

Supporting information: clustering analysis of cell networks

Starting from graphs generated on squared grids of sides of 200 pixel having different number of nodes (25, 50, 100, 150, 200) and connection probability (from 0.1 to 1) combination, we imported the nodes coordinates matrix and the connectivity information (contained in the *adjacency matrix)*, to quantify some network parameters including the clustering coefficient, the characteristic path length and then the 'smallworldness' coefficient. These parameters give an indication of the connectivity properties of the nodes in a graph and allow to distinguish between graphs of different types (regular, random or small world).

In graph theory, the **clustering coefficient (C_c)** 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

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

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, and a *global* value, C_c, is derived upon averaging C_i over all the nodes that compose the graph.

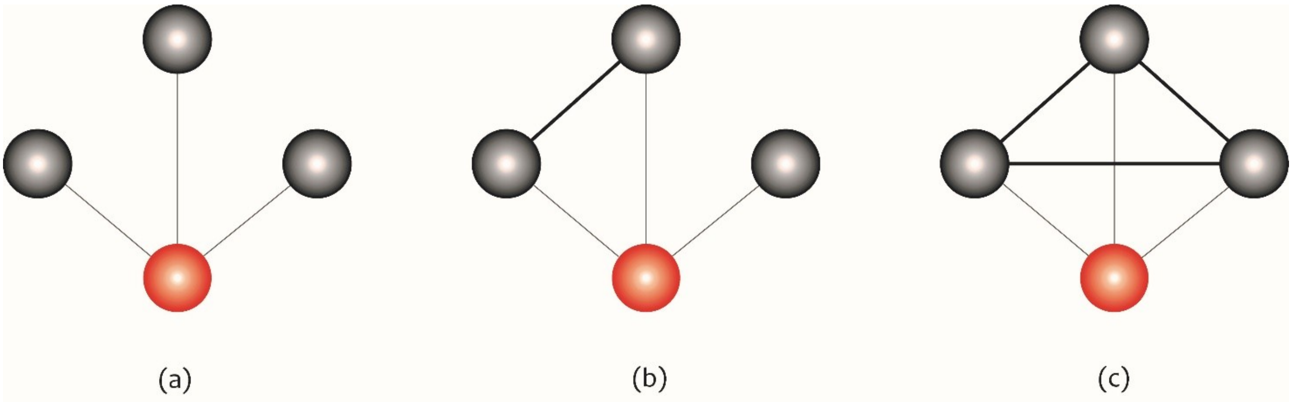

Fig1. Example of three networks and respective clustering coefficients (see Eq.1). (a) There are no edges among the neighbors of node *i* (in red) and $C_i = \frac{2 \cdot 0}{3 \cdot (2)} = 0$; (b) one edge among the neighbors $C_i = \frac{2 \cdot 1}{3 \cdot (2)} = \frac{1}{3}$; (c) three edges among the neighbors $C_i = \frac{2 \cdot 3}{3 \cdot (2)} = 1$.

The **characteristic path length** (Cpl) is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. The *shortest path length* (Spl), between two nodes is the path that connects the two nodes with the shortest number of edges and it is the minimum distance between a generic couple of nodes.

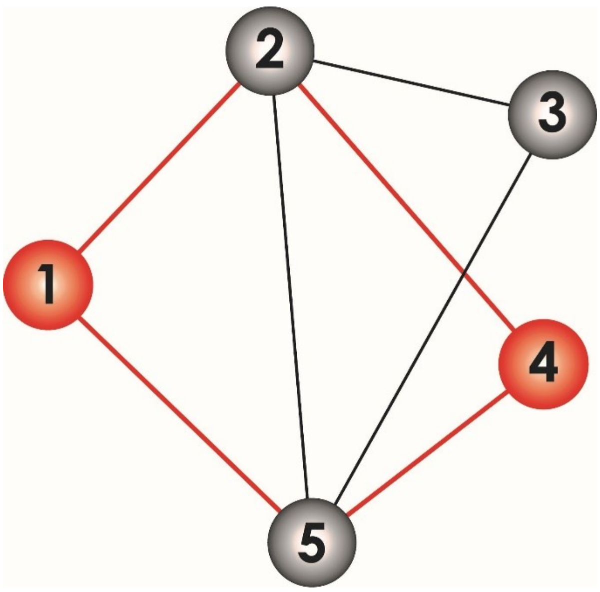

Fig 2. In this example, between nodes 1 and 4, there are two *shortest path* of length 2: {1,2,4} and {1,5,4}. Longer paths between the two nodes are {1, 2, 5, 4}, {1, 5, 2, 4}, {1, 2, 3, 5, 4}, {1, 5, 3, 2, 4}. Shorter paths are desiderable to enhance speed of communication.

We show now how to calculate the Spl for a couple of nodes n_l and n_m premising that, in the imported *adjacency matrix* $A = a_{ij}$, the indices i and j run through the number of nodes n in the graph and $a_{ij} = 1$, if there exists a connection between i and j, $a_{ij} = 0$ otherwise. 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. Notice that this property translates into symmetry of A being $a_{ij} = a_{ji}$. Moreover, $a_{ij} = 0$.
In A, $a_{l,i}$ and $a_{i,m}$ account for all the pairs of nodes which are connected to n_l and n_m respectively. The sum of $a_{l,i}$ and $a_{i,m}$ over all the nodes in A, is stored in a new matrix $A_2 = \sum a_{l,i} a_{i,m}$ for all the l and m and A_2 has the same dimension of A. Now multiplicate A_2 and A repeatedly $A_2 = A_2$A, until all the terms of A_2 are non-zero and those terms in position ij will be the Spl between node i and node j. Finally, the characteristic path length Cpl is calculated like the average of Spl over A_2.

Once obtained the C_c and Cpl values, we defined a precise measure of 'smallworldness', the '**smallworldness' coefficient** (*SW*), based on the trade off between high local clustering and short path length.
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). Let Cpl_u and Cc_u be the mean shortest path length and the mean clustering coefficient for the *E–R* random graphs, obtained meaning 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:

$$\gamma = \frac{Cc_{graph}}{Cc_u} \tag{4}$$

$$\lambda = \frac{Cpl_{graph}}{Cpl_u} \tag{5}$$

Thus, the 'smallworldness' coefficient is

$$SW = \frac{\gamma}{\lambda} \tag{6}$$

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