A Complex Network Approach to the Determination of Functional Groups in the Neural System of C. Elegans

Alex Arenas, Alberto Fernández, and Sergio Gómez

Universitat Rovira i Virgili, Departament d'Enginyeria Informàtica i Matemàtiques, Avinguda dels Països Catalans 26, 43007 Tarragona, Spain
alexandre.arenas@urv.cat
http://deim.urv.cat/~aarenas/

Abstract. The structure of real complex networks is often modular, with sets of nodes more connected between them than to the rest of the network. These communities are usually reflecting a topology-functionality interplay, whose discovery is basic for the understanding of the operation of the networks. Thus, much attention has been driven to the determination of the modular structure of complex networks. Recently it has been shown that this modular organization appears at several scales of description, which may be found by a synchronization process on top of these networks. Here we make use of it for a tentative uncovering of functional groups in the neural system of the nematode C. elegans.

1 Introduction

Complex networks are graphs representative of the intricate connections between elements in many natural and artificial systems [1,2], whose description in terms of statistical properties has been largely developed in the curse for a universal classification of them. However, when the networks are locally analyzed some characteristics that become partially hidden in the statistical description emerge. The most relevant perhaps is the discovery in many of them of *community structure*, meaning the existence of densely (or strongly) connected groups of nodes, with sparse (or weak) connections between these groups [3].

The study of the community structure helps to elucidate the organization of the networks and, eventually, could be related to the functionality of groups of nodes [4]. The most successful solutions to the community detection problem, in terms of accuracy and computational cost required, are those based in the optimization of a quality function called *modularity* and proposed in [5], that allows for the comparison of different partitioning of the network. Given a network partitioned into communities, being C_i the community to which node i is assigned, the mathematical definition of modularity [6] is expressed in terms of the weighted adjacency matrix w_{ij}, that represents the value of the weight of the link between nodes i and j (0 if no link exists), as

$$Q = \frac{1}{2w} \sum_i \sum_j \left(w_{ij} - \frac{w_i w_j}{2w} \right) \delta(C_i, C_j) \ , \tag{1}$$

P. Liò et al. (Eds.): BIOWIRE 2007, LNCS 5151, pp. 9–18, 2008.
© Springer-Verlag Berlin Heidelberg 2008

where the strength of node i is $w_i = \sum_j w_{ij}$, the total strength of the network is $2w = \sum_i w_i$, and the Kronecker delta function $\delta(C_i, C_j)$ takes the value 1 if node i and j are into the same community, 0 otherwise.

The modularity of a given partition is then, up to a multiplicative constant, the probability of having edges falling within groups in the network minus the expected probability in an equivalent (null case) network with the same number of nodes, and edges placed at random preserving the strengths of the nodes. The larger the modularity the best the partitioning is, because more deviates from the null case. Note that the optimization of the modularity cannot be performed by exhaustive search since the number of different partitions are equal to the Bell or exponential numbers [7], which grow at least exponentially in the number of nodes N. Indeed, optimization of modularity is a NP-hard (Non-deterministic Polynomial-time hard) problem [8]. Several authors have attacked the problem proposing different optimization heuristics [9,10,11,12,13,14].

Maximizing modularity one obtains the "best" partition of the network into communities. This partition represents an intermediate topological scale of organization, or *mesoscale*, that in many cases has been shown to coincide with known information about subdivisions in the network [5,15]. However, recently it has been pointed out that the optimization of the modularity has a characteristic scale related to the number of links in the network, that delimits the resolution beyond which no separation into smaller groups can be obtained when optimizing modularity, although these smaller partitions, and then different levels of description, are plausible to exist from direct observation [16]. The problem seems to be that modularity, as it has been prescribed, does not have access to these other levels of description. The reason for this is that the topological scale at which we have access by maximizing modularity has a limit.

Here we propose the use of a synchronization process [17,18,19,20,21] between nodes in the network, for the determination of the mesoscales in complex networks. In particular we show its applicability to the determination of several scales of organization in the synaptic connectivity of the neuronal system of the nematode *C. elegans* from actual data compiled from [22] and arranged by [23].

2 Determination of the Mesoscales

The main idea we propose here to detect the mesoscales is to assimilate all nodes with identical oscillators following the coupling proposed by Kuramoto [24]. The Kuramoto model consists of a population of N coupled phase oscillators where the phase of the ith unit, denoted by $\theta_i(t)$, evolve in time according to the following dynamics

$$\frac{d\theta_i}{dt} = \omega_i + \sum_j K_{ij} sin(\theta_j - \theta_i) \qquad i = 1, ..., N \tag{2}$$

where ω_i stands for the natural frequency of the oscillator and K_{ij} describes the coupling between units. The original model studied by Kuramoto assumed

mean-field interactions $K_{ij} = K,\ \forall i, j$. If the oscillators are identical ($\omega_i = \omega,\ \forall i$) there is only one attractor of the dynamics: the fully synchronized regime where $\theta_i = \theta,\ \forall i$.

The temporal mesoscales of the dynamics of synchronization (of phase oscillators) near the synchronization attractor are governed by the solutions of the linear dynamics:

$$\frac{d\theta_i}{dt} = -k \sum_j L_{ij}\theta_j \qquad i = 1, ..., N \tag{3}$$

where k is a constant, θ_j are the phases of the nodes and L_{ij} the Laplacian matrix of the network, defined as $L_{ij} = w_i\delta_{ij} - w_{ij}$, where w_i is the strength of node i, δ_{ij} is the Kronecker delta and w_{ij} is the element of the weighted adjacency matrix.

To identify patterns of synchronization corresponding to temporal mesoscales, we use a discretization of the matrix $\rho_{ij} = \langle cos(\theta_i - \theta_j)\rangle$ where $\langle\cdots\rangle$ stands for the average over different realizations of the initial conditions. In all cases presented here we have averaged 10^5 realizations [17]. The solution of Eq.3 in terms of the normal modes $\varphi_i(t)$ reads

$$\varphi_i(t) = \sum_j B_{ij}\theta_j = \varphi_i(0)e^{-\lambda_i t} \qquad i = 1, ..., N \tag{4}$$

where λ_i are the eigenvalues of the Laplacian matrix, and B is the matrix of its eigenvectors. The different intermediate scales are separated according to gaps in the mode decay times defined by the difference between the inverse of consecutive (ordered) eigenvalues $1/\lambda_i - 1/\lambda_{i+1}$. Note that the smallest (different from 0) eigenvalue of the Laplacian matrix determines the time scale for the whole network to synchronize.

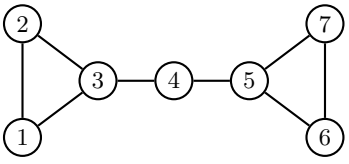
By using this method it is possible to reveal the structural mesoscales following the synchronization process, from the beginning of the process (t=0) when oscillators are desynchronized, to the end of the process when the whole system is at the synchronization fixed point. The mesoscales are defined as the communities of synchronization by taking snapshots of the evolution at discrete times and finding the synchronization patterns as described above.

3 Analysis of the Mesoscales

The different structural patterns observed along the synchronization process can be represented as a set of matrices at different times. For a comprehensive representation of the whole mesoscale that allows for the extraction of information, we propose to represent each matrix after processing it as follows: (i) for each pair of nodes we compute the degree of synchronization; (ii) the matrix is reordered from left to right by the size of the connected components with larger synchronization values. The darker colors in the scale represent groups of nodes that are more synchronized.

In the following subsections we give the details of the mesoscales determination for a toy model, in order to clarify all the steps involved.

A



B

patterns found	% length
$\{1,2,3,4\}$ $\{5,6,7\}$	58.25%
$\{1,2,3\}$ $\{4\}$ $\{5,6,7\}$	36.52%
$\{1,2\}$ $\{3\}$ $\{4\}$ $\{5\}$ $\{6,7\}$	5.23%

C

	1	2	3	4	5	6	7
1	1.00	1.00	0.95	0.58	0.00	0.00	0.00
2	1.00	1.00	0.95	0.58	0.00	0.00	0.00
3	0.95	0.95	1.00	0.58	0.00	0.00	0.00
4	0.58	0.58	0.58	1.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	1.00	0.95	0.95
6	0.00	0.00	0.00	0.00	0.95	1.00	1.00
7	0.00	0.00	0.00	0.00	0.95	1.00	1.00

D

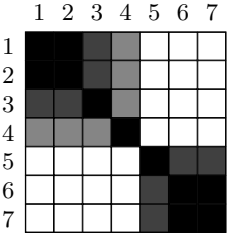


Fig. 1. (A) Sample network for the determination of its mesoscales. **(B)** Lengths corresponding to each optimal configuration. **(C)** Mesoscales table, formed by the lengths of pairs of nodes in the same community, normalized by the total length. **(D)** Mesoscales matrix (the contrast has been adjusted to enhance the visibility of the four different length levels present in the mesoscales table).

3.1 Mesoscales Matrix

Let us consider the undirected graph in Fig. 1A, with all weights equal to 1. We study the mesoscales using a discretization of the synchronization process. Any graphical representation of the whole temporal mesoscale should take into account, for every pair of nodes, the proportion of mesoscales at which they belong to the same community. Each mesoscale has a natural *length* (see Fig. 1B) defined by the range of values of time (in logarithmic scale) at which the patterns are represented. Thus, the length proportion for a pair of nodes is the sum of the lengths corresponding to mesoscales in which they belong to the same community, normalized by the total length (see Fig. 1C). The graphical representation of this table, which we call *mesoscales matrix*, is shown in Fig. 1D.

3.2 Filtered Mesoscales Matrix

The previous example is quite simple since the mesoscales obtained are hierarchical and their representation following the hierarchical order is convenient

A

	1	2	3	4	5	6	7
1	1.0	0.2	0.1	0.8	0.0	0.0	0.1
2	0.2	1.0	0.4	0.0	0.6	0.0	0.8
3	0.1	0.4	1.0	0.2	0.1	0.7	0.1
4	0.8	0.0	0.2	1.0	0.2	0.1	0.0
5	0.0	0.6	0.1	0.2	1.0	0.0	0.1
6	0.0	0.0	0.7	0.1	0.0	1.0	0.3
7	0.1	0.8	0.1	0.0	0.1	0.3	1.0

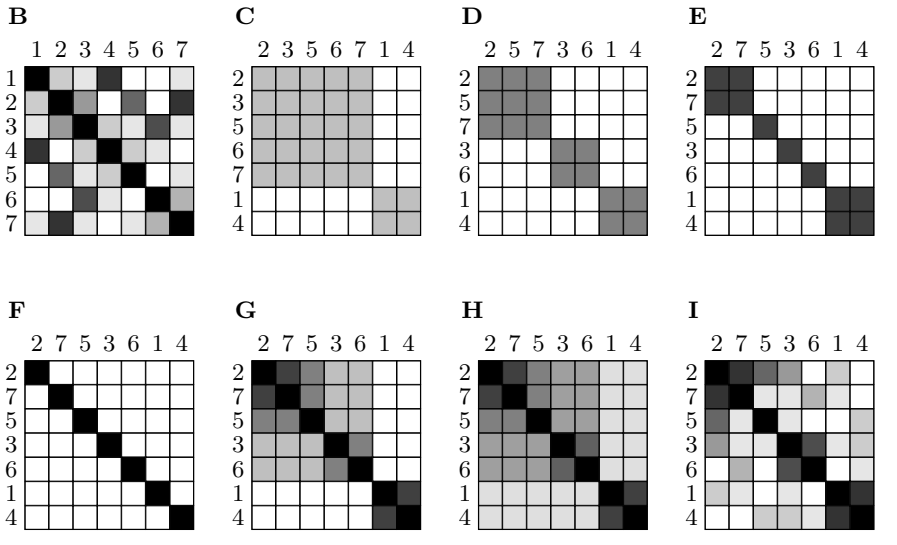


Fig. 2. Determination of the filtered mesoscales matrix. **(A)** Sample mesoscales table. **(B)** Corresponding mesoscales matrix. **(C)** Connected components of the mesoscales matrix at the threshold of 0.25. **(D)** Threshold of 0.50. **(E)** Threshold of 0.75. **(F)** Threshold of 1.00. **(G)** Filtered mesoscales matrix (4 levels). **(H)** Filtered mesoscales matrix (8 levels). **(I)** Mesoscales matrix using the ordering defined by the filtered mesoscales matrix.

to extract information. However, let us suppose that, after averaging, we have obtained the mesoscales table in Fig. 2A, whose mesoscales matrix is shown in Fig. 2B. We define the *filtered mesoscales matrix* which is obtained by the application of several thresholds to the mesoscales matrix, i.e. the lengths below the threshold are discarded, and the connected components of the graph defined by the remaining lengths are found. Figures 2C–F show the results after the application of thresholds 0.25, 0.50, 0.75 and 1.00. The first threshold divides the network in two connected components, which ordered by size are: $\{2, 3, 5, 6, 7\}$, $\{1, 4\}$. This partition gives the reference for the rest of the process. The following connected components, ordered by size within each one of the groups found in the previous threshold, are: $\{2, 5, 7\}$, $\{3, 6\}$, $\{1, 4\}$ for threshold 0.50; $\{2, 7\}$,

$\{5\}, \{3\}, \{6\}, \{1, 4\}$ for threshold 0.75; and $\{2\}, \{7\}, \{5\}, \{3\}, \{6\}, \{1\}, \{4\}$ for threshold 1.00. Finally, the filtered mesoscales matrix is built by the composition of these four threshold matrices (see Fig. 2G).

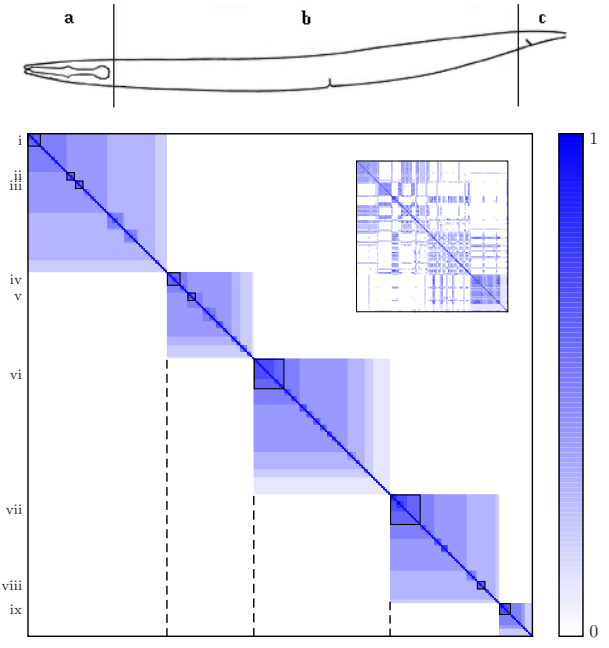
In order to complete this example, we give two more matrices. First, we want to show that by using more threshold cuts in the mesoscales matrix we would obtain a more detailed filtered mesoscales matrix preserving the structures already found because of transitivity. For instance, the result using eight instead of four thresholds is given in Fig. 2H. Second, we would like to assert the difference between the mesoscales matrix and the filtered mesoscales matrix. For this reason we show in Fig. 2I the former using the ordering found by the latter. Clearly, the definition of the filtered mesoscales matrix helps to extract information of the mesoscales imposing transitivity relations in the data found by the mesoscales matrix.

4 Analysis of the C. Elegans Neuronal Network

Here we develop the analysis of the C. elegans neuronal network. We have taken the largest connected component of the C. elegans neuronal network (297 neurons), and analyzed the synchronization dynamics in 1000 exponentially distributed time steps up to complete synchronization.

The neuronal network of the C. elegans can be represented as a weighted adjacency matrix. The order of the neurons in the matrix follows that of [1], obtained from experimental data in [22]. The detection of the mesoscales in this neuronal system has been performed according to the method explained in this paper. The best partition corresponding to the Newman's modularity definition provides with 5 communities. They all contain neurons whose soma can be correlated with spatial parts of the worm, mainly the head, the body and the tail (see Fig. 3). This coarse graining provides then with a large scale structural level in this system. We use the Newman's partition as a reference for the substructures found by the method, i.e. Newman's scale corresponds to the threshold equal to 0 in the mesoscales matrix.

Any trial of classification by the functional role of neurons in the C. elegans is extremely delicate because of the multi functional aspects they have. Many neurons participate in different synaptic pathways resulting in different functionalities. To extract information from the results obtained, we use the filtered mesoscales matrix as explained in the previous section. By fixing a threshold in the length value, we are able to unravel sub-structural scales that could correspond to groups of neurons involved in different functionalities. The most interesting information is that provided at a large value of the threshold, because in this case the substructures found contain small groups of neurons whose activity is most likely associated to a specific action. With this information at hand, and the wide description of each neuron found at the public database of C. elegans [22], we propose a tentative classification of some groups of neurons by functionality.



a	96.4 %	92.0 %	15.2 %	40.6 %	36.8 %
b	1.2 %	0.0 %	82.3 %	32.8 %	26.4 %
c	2.4 %	8.0 %	2.5 %	26.6 %	36.8 %

Fig. 3. Mesoscales matrix (inside) and filtered mesoscales matrix of the C. elegans neuronal network

We have studied the filtered mesoscales matrix at a threshold value of 0.6. Fixing our attention at this level of description, we present a tentative functional classification for the groups of five or more neurons (see Fig. 3). We have used the information presented in [25] and [22] for each neuron position and individual functionality, as a guide for the classification of specific actions. Our purpose, after identification of individual functionalities, has been to assign a specific action to the whole group of neurons.

The results of the analysis of the filtered mesoscales matrix for the *C. elegans* neuronal connectivity show that: i) the substructures that prevail at different topological scales are most of them in agreement with the location of the soma of neurons along the body of the worm, and ii) the functionality of the different substructures found by the method are correlated with specific actions of the worm which allows for a tentative classification of functional groups. The classification obtained (see Table 1) does not pretend to be exact but to provide biologists with a useful information for future research.

Table 1. Tentative functionalities of the groups of five or more neurons of the C. elegans at a threshold level of 0.6 in the filtered mesoscales matrix

	Neurons	Tentative function
i	RIAL, RIAR, RMDR, RMDVR, SMDVL, SMDVR, RMDDL, SMDDR	Nose/head orientation movement.
ii	IL1DR, IL1VR, IL2DR, IL2VR, RIPR	Head-withdrawal reflex, more related to dorsal relaxation. When worms are touched on either the dorsal or ventral sides of their nose with an eyelash, they interrupt the normal pattern of foraging and undergo an aversive head-withdrawal reflex.
iii	IL2L, IL2R, OLQVL, OLQVR, RIH	Head-withdrawal reflex, more related to ventral relaxation.
iv	ADLR, AIBR, ASEL, ASHR, AWCL, AWCR, AIAR, AIYL	Olfactory and thermo sensation reflex.
v	ASGL, ASJL, ASKL, AIAL, PVQL	Chemotaxis to lysine reflex.
vi	DB1, DB2, DD1, VB2, VD2, AS3, DA2, DA3, DA4, DA5, DB3, DB4, VA3, VD3, VD4, VD5, VD6, WM	Backward/sinusoidal movement of the worm, more related to touch stimulus.
vii	AVAL, AVAR, AVBL, AVBR, AVDL, AVDR, AVEL, AVER, DA1, FLPL, FLPR, RIFR, PVDL, PVDR, PVPR, PQR, PVCL, PVCR	Forward and backward/sinusoidal movement of the worm, more related to search for food in starving case, involve social feeding effect.
viii	AVHL, AVHR, AVJL, AVFL, AVFR	Impossible to determine from the experimental data available. There is not any specific function known for any of these neurons.
ix	AVKL, AVKR, PDEL, PDER, PVM, DVA, WN	The functionality of this group could be related to a relaxation state similar to a sleep state, with reduced motor activity, decreased sensory threshold, characteristic posture and easy reversibility, basically mediated by PDs neurons.

5 Conclusions

In this paper we have introduced a method to uncover information from the several scales of description found in many real complex systems. The result is a *mesoscales matrix* whose representation provides with a structural map of the topology of the network. The mesoscale matrix for the sample network presented reveals how nodes form groups at different scales. Nevertheless, the symmetries of the network play in favor of this clear visualization. In real complex networks, where these symmetries are usually absent, a filtering process is needed to reveal the same information.

Hence, we have also designed what we call the *filtered mesoscales matrix*, consisting in to: (i) fix a mesoscale (a level color) and remove from the mesoscales

matrix the elements under this level (lighter colors), (ii) calculate the connected components of the remaining elements (groups), and (iii) reorder the matrix from left to right in decreasing size within the groups obtained at previous levels. This process is iterated starting from the lowest mesoscale to the highest one, accumulating the results of previous stages. This way, without losing any information from the original mesoscales matrix, we achieve a clearer representation of the structural map.

We have applied the complete method to unravel the mesoscales of the neuronal connectivity of the nematode *C. elegans*. The whole nervous system of the nematode can be represented as a weighted adjacency matrix. We have calculated the corresponding filtered mesoscales matrix. The results of the analysis of the filtered mesoscales matrix for the *C. elegans* show some interesting correlations between synchronization patterns with the location of the soma of neurons, and with the functionalities in the worm. These results could help biologists to design specific targeted experiments based on the classification of neurons according to their roles at different topological scales.

Acknowledgments. This work has been supported by Spanish Ministry of Science and Technology Grant FIS2006-13321-C02-02.

References

1. Strogatz, S.H.: Exploring complex networks. Nature 410, 268–276 (2001)
2. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U.: Complex networks: structure and dynamics. Phys. Rep. 424, 175–308 (2006)
3. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. USA 99, 7821–7826 (2002)
4. Guimerà, R., Amaral, L.A.N.: Functional cartography of metabolic networks. Nature 433, 895–900 (2005a)
5. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113 (2004)
6. Newman, M.E.J.: Analysis of weighted networks. Phys. Rev. E 70, 056131 (2004a)
7. Bell, E.T.: Exponential Numbers. Amer. Math. Monthly 41, 411–419 (1934)
8. Brandes, U., Delling, D., Gaertler, M., Goerke, R., Hoefer, M., Nikoloski, Z., Wagner, D.: Maximizing Modularity is hard. arXiv:physics/0608255 (2006)
9. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Phys. Rev. E 70, 066111 (2004)
10. Duch, J., Arenas, A.: Community identification using Extremal Optimization. Phys. Rev. E 72, 027104 (2005)
11. Guimerà, R., Amaral, L.A.N.: Cartography of complex networks: modules and universal roles. J. Stat. Mech., P02001 (2005b)
12. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. Phys. Rev. E 69, 066133 (2004b)
13. Newman, M.E.J.: Modularity and community structure in networks. Proc. Natl. Acad. Sci. USA 103, 8577–8582 (2006)
14. Pujol, J.M., Béjar, J., Delgado, J.: Clustering Algorithm for Determining Community Structure in Large Networks. Phys. Rev. E 74, 016107 (2006)

15. Danon, L., Díaz-Guilera, A., Duch, J., Arenas, A.: Community analysis in social networks. J. Stat. Mech., P09008 (2005)
16. Fortunato, S., Barthélemy, M.: Resolution limit in community detection. Proc. Natl. Acad. Sci. USA 104, 36–41 (2007)
17. Arenas, A., Díaz-Guilera, A., Perez-Vicente, C.J.: Synchronization reveals topological scales in complex networks. Phys. Rev. Lett. 96, 114102 (2006)
18. Arenas, A., Díaz-Guilera, A., Perez-Vicente, C.J.: Synchronization processes in complex networks Physica D 224, 27–34 (2006)
19. Gómez-Gardeñes, J., Moreno, Y., Arenas, A.: Paths to synchronization on complex networks. Phys. Rev. Lett. 98, 034101 (2007)
20. Gómez-Gardeñes, J., Moreno, Y., Arenas, A.: Synchronizability determined by coupling strengths and topology on Complex Networks. Phys. Rev. E 75, 066106 (2007)
21. Boccaletti, S., Ivanchenko, M., Latora, V., Pluchino, A., Rapisarda, A.: Detecting complex network modularity by dynamical clustering. Phys. Rev. E 75, 045102(R) (2007)
22. White, J.G., Southgate, E., Thompson, J.N., Brenner, S.: The structure of the nervous system of the nematode caenorhabditis elegans. Phil. Trans. Royal Soc. London. Series B 314, 1–340 (1986)
23. Achacoso, T.B., Yamamoto, W.S.: AY's Neuroanatomy of C. Elegans for Computation. CRC Press, Boca Raton (1992)
24. Kuramoto, Y.: Self-entrainment of a population of coupled nonlinear oscillators. Lect. Notes in Physics 30, 420–422 (1975)
25. Durbin, R.M.: Studies on the Development and Organisation of the Nervous System of Caenorhabditis elegans. PhD Thesis, University of Cambridge (1987)