2016 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2016, Yugawara, Japan, November 27th-30th, 2016

Large Graph Laplacian Matrix and Functional Map of Whole Brain of *C. elegans*

Hiromu Sakuma†, Takayuki Teramoto‡, Sayuri Kuge‡, Takeshi Ishihara‡ and Yuishi Iwasaki†

†Graduate School of Science and Engineering, Ibaraki University
4–12–1 Nakanarusawa, Hitachi, Ibaraki 316-8511, Japan
‡Graduate School of Science, Kyushu University
744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan
15nm911r@vc.ibaraki.ac.jp, teramoto.takayuki.873@m.kyushu-u.ac.jp,
kuge.sayuri.449@m.kyushu-u.ac.jp, takeiscb@kyushu-u.org,
yuishi.iwasaki.sc@vc.ibaraki.ac.jp

Abstract—Towards understanding of neural signaling of *Caenorhabditis elegans*, cluster analysis is carried out for the central whole brain imaging data on the basis of spectral clustering. Correlation between the data is represented as weighted edges in a similarity graph. Optimized clustering resolves into an eigenvalue problem of the graph Laplacian defined by the similarity graph. We analyze the neural activities of the wild-type and the *unc*-7 mutant which has defect in gap junction. In the wild-type, the neural activities within the same cluster are highly coherent. There are anti-phase clusters in which the neural activities are obviously in anti-phase. In the *unc*-7 mutant, highly coherent neural activities are disappeared. Gap junction is required to generate a highly organized neural synchronization. In functional maps of the neurons, the functional left-right symmetry of neuronal position within the same cluster is partially shown.

1. Introduction

The nematode *Caenorhabditis elegans* is subject to the brain activity map project together with *Drosophila*, zebrafish and mouse [1]. *C. elegans* is a useful model organism in neurobiology because the neural connectivity is fully known [2]. The nervous system of *C. elegans* consists of 302 neurons connected by about 6500 chemical synapses and about 900 gap junctions. About 170 neurons of them densely locate in the head region. Although the wiring diagram of the nervous system is determined, functional map is poorly known. Recently, we developed a 4D imaging system to measure the neural activity in the head region as a worm lives [3]. Therefore, cluster analysis is carried out for the central whole brain imaging data.

The *k*-means algorithm and the hierarchical clustering are popular clustering methods. Since the *k*-means algorithm has a tendency to divide objects into equally sized clusters, an obviously inappropriate partition is frequently occurred. In the hierarchical clustering, objects tend to be added to the tail of the largest cluster. This is known as "chaining phenomenon" and often generates spurious clusters. In this work, spectral clustering [4] is applied to determine appropriate clusters of the neural imaging data. Correlation between the data is represented as weighted edges in a similarity graph. The graph Laplacian is subsequently defined by the similarity graph. Spectral clustering finds group structure in a given data set on the basis of the eigenvectors of the graph Laplacian.

2. Method

2.1. Graph Laplacian

For a given graph with N vertices, its $N \times N$ graph Laplacian L is defined by

$$L = D - W, \tag{1}$$

where D and W are the $N \times N$ degree matrix and the $N \times N$ adjacency matrix, respectively. The element w_{ij} of W corresponds to an edge for a directed graph or an arc for an undirected graph whose value represents the weight from i-th vertex to j-th vertex. The element of D is $d_{ii} = \sum_{j=1, j\neq i}^{N} w_{ij}$ or $d_{ij} = 0$ for $i \neq j$.

There are several important properties in the graph Laplacian. Let λ_k and $\vec{\xi}_k$ be the eigenvalue and the eigenvector of L, respectively.

$$L\vec{\xi}_k = \lambda_k \vec{\xi}_k. \tag{2}$$

When w_{ij} is non-negative and symmetric, that is $w_{ij} = w_{ji} \geq 0$, the eigenvalues are real and non-negative. The smallest eigenvalue is 0 and its eigenvector is $\vec{1} = (1, 1, \cdots, 1)$. When the eigenvalues are arranged in an ascending order, therefore, $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$. The multiplicity M of $\lambda_k = 0$ ($k = 1, \cdots, M$) is equal to the number of connected components $G_1, \cdots, G_M$ in the graph. The eigenspace of $\lambda_k = 0$ is spanned by the vectors $\vec{1}_{G_1}, \cdots, \vec{1}_{G_M}$ of those components. Spectral clustering is indebted to these properties.

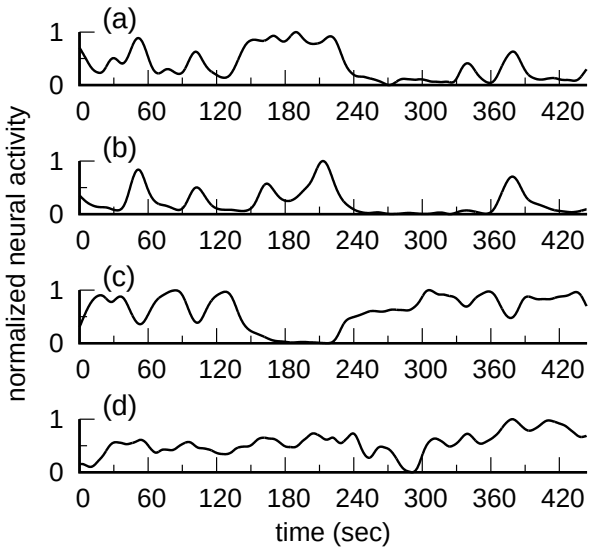

Figure 1: The neural activity measured in the different neurons of the wild-type. The Pearson's correlation coefficient r_{ij} between the data (a) and (b,c,d) are $r_{\rm ab} = 0.757$, $r_{\rm ac} = -0.860$ and $r_{\rm ad} = 0.129$, respectively. The correlation between the data (b) and (c) is large negative, $r_{\rm bc} = -0.662$.

2.2. Spectral clustering

First of all, a similarity graph in which each vertex represents each data is considered. A certain degree of "similarity" between the data, whose value w_{ij} is non-negative and symmetric, is attached as an edge so that the similarity graph is undirected. In this work, the similarity graph is divided into K clusters $G_1, \cdots, G_K$ whose partition satisfies $G_k \cap G_{k'} = 0$ for $k \neq k'$ and minimizes the following cost function [5].

$$\mathrm{NCut}(G_1, \cdots, G_K) = \sum_{k=1}^{K} \frac{\mathrm{cut}(G_k, \bar{G_k})}{\mathrm{vol}(G_k)}. \qquad (3)$$

Where $\mathrm{cut}(A, B) = \sum_{i \in A, j \in B} w_{ij}/2$, $\bar{A}$ is the complement of A and $\mathrm{vol}(G_k) = \sum_{i \in G_k} d_{ii}$. The number of clusters K is a given number. The cost function in eq. (3) indicates the sum of the edges w_{ij} between G_k and $\bar{G_k}$ normalized by the factor $\mathrm{vol}(G_k)$.

Since this optimization problem is NP-hard [6], it is difficult to determine optimal clustering for a large graph. In the case of nervous system of *C. elegans*, for examples, the number of vertices is $N \approx 170$ for the head region or $N = 302$ for the entire nervous system. According to the Rayleigh-Ritz theorem, the cost function in eq. (3) is approximately replaced with the sum of the eigenvalues λ'_k of the normalized graph Laplacian $L_{\rm sym} = D^{-1/2} L D^{-1/2}$ for optimization.

$$\min_{G_1, \cdots, G_K} \mathrm{NCut}(G_1, \cdots, G_K) \to \min_{\lambda'_1, \cdots, \lambda'_K} \sum_{k=1}^{K} \lambda'_k. \qquad (4)$$

Practically, the optimization problem resolves into an eigenvalue problem of $L_{\rm sym}$. Spectral clustering algorithm is as follows [4].

step 1. First of all, W of the similarity graph is given. Subsequently, D, L and $L_{\rm sym}$ are defined from W.

step 2. The eigenvalues λ'_k and the eigenvectors $\vec{\xi}'_k$ of $L_{\rm sym}$ are calculated: $L_{\rm sym}\vec{\xi}'_k = \lambda'_k \vec{\xi}'_k$.

step 3. Using $\vec{\xi}'_k$ corresponding to the smallest K eigenvalues, $N \times K$ matrix $U = (u_{i,k})$ is introduced. Where $u_{i,k} = \xi'_{i,k} / \sqrt{\sum_{j=1}^{K} (\xi'_{i,j})^2}$. Let $\vec{v}_i$ be the K dimensional vector corresponding to i-th row of U. The points $(v_i)_{i=1,\cdots,N}$ are divided into K clusters $C_1, \cdots, C_K$ by the k-means algorithm.

In this work, W is given by the Pearson's correlation coefficient r_{ij} between i-th and j-th neuron data. Since w_{ij} must be non-negative in a similarity graph, the absolute value of r_{ij} is introduced to define the graph Laplacian, that is $w_{ij} = |r_{ij}|$. A negative correlation works in the same way as a positive correlation in the spectral clustering. For examples, the data in Fig.1a, Fig.1b and Fig.1c probably belong to the same cluster although the correlation coefficient $r_{\rm bc} = -0.662$ is large negative. The sign of r_{ij} is not taken into account.

All programs in this work are written in the C language and are compiled by the GNU C compiler on UNIX.

2.3. Experimental data

For the neurons in the head region of *C. elegans*, somatic Ca^{2+} concentrations are simultaneously visualized by the 4D live-cell imaging system [3]. Voltage-gated Na^+ channels have not been found in *C. elegans* [7]. Instead of Na^+-based classical action potentials, the neurons might have Ca^{2+}-based signal amplification as in the large nematode *Ascaris*. Therefore, the Ca^{2+} signaling well reflects the dynamics of the neural activity. Using the Ca^{2+}-binding fluorescent protein YC2.6, Ca^{2+} imaging data is measured for the wild-type (the standard "normal" type) and the *unc*-7 mutant. The *unc*-7 mutant has defect in gap junction and exhibits uncoordinated locomotion. The neural connectivity of the *unc*-7 mutant is different from that of the wild-type. No stimulation is add to a worm during measurement. Time interval is 1/4.5 sec. In this work, we adopt the time course data sets of 2000 time points, $x_i(n)$ ($n = 1, 2, \cdots, 2000$). Here i and n are neuron and time indices, respectively. The number of observed neurons is about $N \approx 150$. To eliminate the YC2.6 expression difference in each neuron, we analyze the normalized values $(x_i(n) - x_{\min,i})/(x_{\max,i} - x_{\min,i})$. $x_{\max,i}$ and $x_{\min,i}$ are the maximum and the minimum values of $x_i(n)$, respectively. Examples of the neural activity are shown in Fig.1.

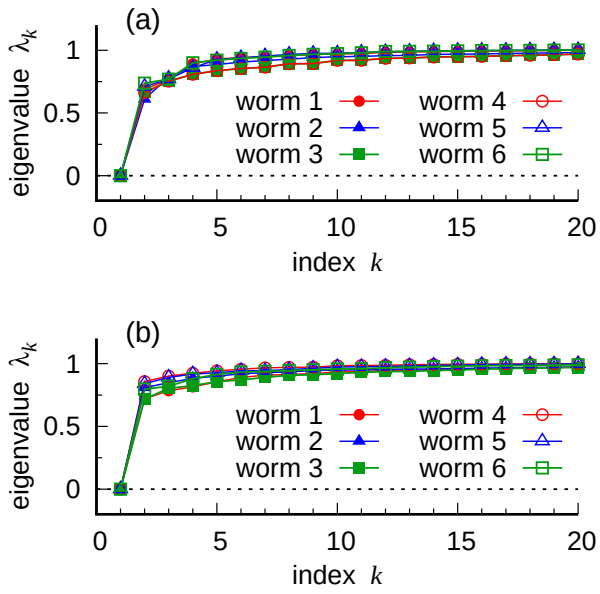

Figure 2: Spectra of the normalized graph Laplacian L_{sym} for the wild-type (a) and the *unc*-7 mutant (b). The smallest 20 eigenvalues are plotted in an ascending order for 6 worms.

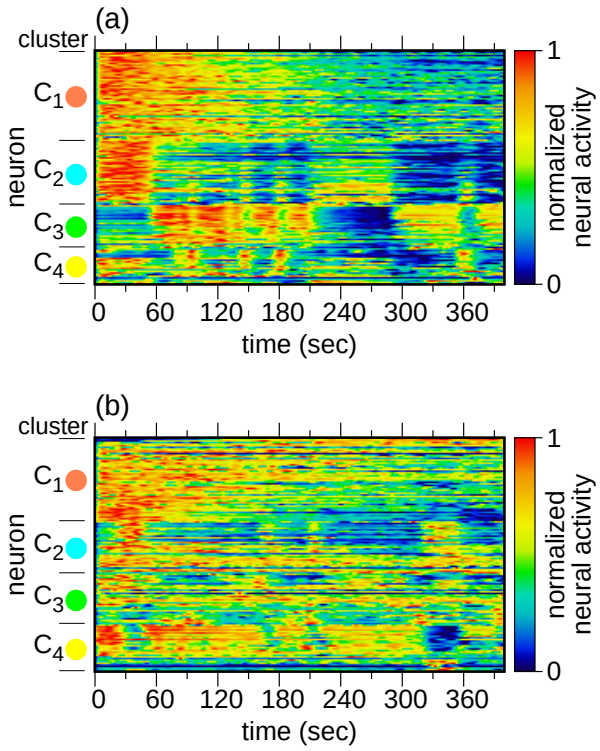

Figure 3: Spectral clustering of the neural activity for the wild-type (a) and the *unc*-7 mutant (b). The neural activity is displayed by heatmap. The smallest K eigenvalues of L_{sym} in Fig.2 are used for clustering. $K = 4$.

3. Results

3.1. Spectrum of graph Laplacian

Spectra of the normalized graph Laplacian L_{sym} are shown in Fig.2. We confirm that the smallest eigenvalue of the graph Laplacian is $\lambda_1 = 0$. Eq. (4) indicates that the $(K+1)$-th eigenvalue λ_{K+1} gives the sum of newly cut weights w_{ij}, that is an additional "cost", when the number of partitions is increased from K to $(K+1)$. At a large spectral gap, therefore, no loosely connected cluster remains. The graph is already divided into strongly connected clusters. Since a large spectral gap is not shown in Fig.2, an appropriate K is not determined to indicate a distinct division of the graph. The graph is connected via the edge w_{ij} which takes a continuous value between 0 and 1. Therefore, the spectra of the graph Laplacian become continuous.

When W is binarized using the threshold ε such as $w_{ij} = 1$ for strong correlation $|r_{ij}| \geq \varepsilon$ or $w_{ij} = 0$ for weak correlation $|r_{ij}| < \varepsilon$, the spectrum of the graph Laplacian has a large gap. When $\varepsilon = 0.4$, for an example, the graph Laplacian has a spectral gap at $k = 4 \sim 7$ (results not shown in this paper).

3.2. Functional map of neurons

In the case of $K = 4$, the spectral clustering of the central whole brain imaging data are shown in Fig.3. In the wild-type (Fig.3a), the neural activities within the same cluster are highly coherent. The clusters C_2 and C_3 are in anti-phase. The clusters C_2 and C_4 are almost in phase. There are many connections among C_1, C_2 and C_3. On the other hand, C_4 has a few connections between the other clusters. As the number of partitions K increases, the clusters are divided into small clusters in which the neural activities are coherent. Correlation between the clusters are consistent with that of the synaptic connectivity in *C. elegans* (results not shown in this paper).

In the *unc*-7 mutant (Fig.3b), highly coherent neural activities are disappeared. The clusters C_2 and C_4 are in anti-phase. The cluster C_3 has a few connections between the other clusters. In addition to the *unc*-7 mutant, we analyze the *unc*-13 mutant which has defect in chemical synapse. In the *unc*-13 mutant, highly coherent neural activities are also disappeared (results not shown in this paper). We find that both gap junction and chemical synapse are required to generate a highly organized neural synchronization.

Functional maps of the neurons are shown in Fig.4. The neurons in the same cluster are filled with the same color and are located at the measurement point in a worm. The neural activity is measured in soma. Each neuron extends long axons from its soma and is connected to other neurons at end point of the axon. Therefore, a physical distance between the somas is not related to a "similarity distance" between the neural activities in the somas. In Fig.4, the neurons in the same cluster are not localized spatially. In

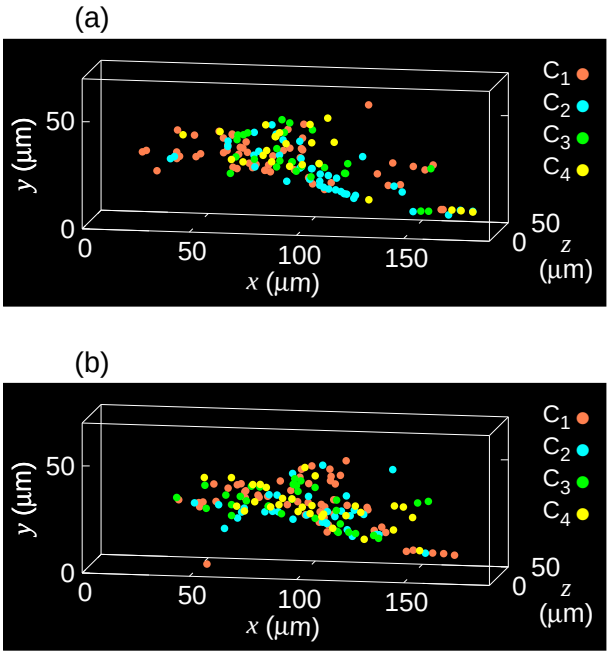

Figure 4: Functional map of the neurons for the wild-type (a) and the *unc*-7 mutant (b). The neurons in the head region are plotted according to the clustering in Fig.3. The *x*, *y* and *z* axes correspond to the the head/tail, the dorsal /ventral, and left/right directions, respectively.

C. elegans, many neurons, AVAL and AVAR for examples, are in pairs and locate on the left and right sides of a worm. Within the same cluster, this left-right symmetry of neuronal position is partially shown in Fig.4.

4. Discussion

Spectral clustering using the normalized graph Laplacian is applied to divide the neural activities in *C. elegans*. Instead of eq. (3), the following cost function is considered [8].

$$\mathrm{RatioCut}(G_1, \cdots, G_K) = \sum_{k=1}^{K} \frac{\mathrm{cut}(G_k, \bar{G_k})}{|G_k|}, \qquad (5)$$

where $|A|$ is the number of vertices in A. The spectral clustering with eq. (5) is achieved by solving the eigenvalue problem of the graph Laplacian L in eq. (2). In this case, spectral clustering divides the neurons into a large cluster and other tiny clusters for the wild-type and the *unc*-7 mutant. As the number of partitions K is increased, an additionally divided cluster is tiny.

To provide a large spectral gap of the graph Laplacian, W of the similarity graph must be sparse. W represents a neighborhood graph in which the "thin" edges are removed. Although the correlation coefficient is used to determine W in tis work, binarized graph Laplacian L is able to be determined directly from the neural data [9].

Since individual difference in the neuron distribution is quite large, neuron annotation, that is a one-to-one correspondence between the imaging data and neuron name, is poorly succeed. When the annotation is completed, the presented cluster analysis is helpful to understand neural signaling of *C. elegans*.

Acknowledgments

The authors would like to thank Y. Iino, Y. Toyoshima, T. Tokunaga, O. Hirose and R. Yoshida for valuable discussions on Ca^{2+} imaging, and K. Takeda for helpful suggestions on spectral clustering. This work was supported by CREST program of Japan Science and Technology Agency.

References

[1] A. P. Alivisatos, M. Chun, G. M. Church, R. J. Greenspan, M. L. Roukes and R. Yuste, "The brain activity map project and the challenge of functional connectomics," *Neuron*, vol.74, pp.970–974, 2012.

[2] J. G. White, E. Soughgate, J. N. Thomson and S. Brenner, "The structure of the nervous system of the nematode *Caenorhabditis elegans*," *Phil. Trans. R. Soc. London, B, Biol. Sci.*, vol.314, pp.1–340, 1986.

[3] T. Tokunaga, O. Hirose, S. Kawaguchi, Y. Toyoshima, T. Teramoto, H. Ikebata, S. Kuge, T. Ishihara, Y. Iino and R. Yoshida, "Automated detection and tracking of many cells by using 4D live-cell imaging data," *Bioinformatics*, vol.30, pp.i43–i51, 2014.

[4] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol.17, pp,395–416, 2007.

[5] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Patt. Analysis Machine Intelligence*, vol.22, pp.888–905, 2000.

[6] D. Wagner and F. Wagner, "Between min cut and graph bisection," in *Proc. 18th Int. Symp. Math. Foundations of Computer Science*, pp.744–750, 1993.

[7] M. B. Goodman, D. H. Hall, L. Avery and S. R. Lockery, "Active currents regulate sensitivity and dynamic range in *C. elegans* neurons," *Neuron*, vol.20, pp.763–772, 1998.

[8] L. Hagen and A. B. Kahng, "New spectral methods for ratio cut partitioning and clustering," *IEEE Trans. Computer-Aided Design*, vol.11, pp.1074–1085, 1992.

[9] H. Sakuma, T. Teramoto, S. Kuge, T. Tokunaga, O. Hirose, Y. Toyoshima, Y. Iino, R. Yoshida, T. Ishihara and Y. Iwasaki, "Intrinsic graph structure estimation from whole-brain imaging data of *C. elegans*," in *Proc. 16th Int. Conf. Systems Biol.*, Poster No.159, 2015.