Global Structure of Directed Networks Emerging from a Category Theoretical Formulation of the Idea “Objects as Processes, Interactions as Interfaces”

Taichi Haruna[1,2]

[1]Department of Earth & Planetary Sciences, Graduate School of Science, Kobe University,
1-1, Rokkodaicho, Nada, Kobe, 657-8501, Japan
[2]PRESTO, Japan Science and Technology Agency (JST), 4-1-8 Honcho Kawaguchi, Saitama, Japan
tharuna@penguin.kobe-u.ac.jp

Abstract

A system of interacting elements can be represented by a directed network so that elements are nodes and interaction between two elements is an arc. Conventionally, each node is just a point, each arc represents some kind of interaction between two nodes and nothing more after the system is mapped to a directed network. However, in many real systems, each element has its own intra-node process and interaction between two elements can be seen as an interface between two intra-node processes. We can formalize this idea “objects as processes, interactions as interfaces” within the framework of category theory. We show that a new notion of connectedness called *lateral connectedness* emerges as a canonical structure obtained from the idea. Lateral connectedness is not defined on the set of nodes of a directed network, but on the set of arcs. By its definition, it may be associated with functional commonality between arcs emerging from shared input or output. As a first application, we examine significance of lateral connectedness in the neuronal network of the nematode *Caenorhabditis elegans* by comparing the partition of the set of arcs induced by the connectedness and the partitions based on neuron functions. Lateral connectedness can capture a part of functional segregation of the neuronal network above a certain interaction strength level.

Introduction

Science of complex networks is one promising approach to understand the intrinsic organization of living systems (Alon, 2006; Junker and Schreiber, 2008; Sporns, 2011). Many characteristics such as degree distributions, average path length, clustering coefficients, centralities, assortativity coefficient, network motifs have been introduced in order to reveal functionality of biological, social, technological systems from network topology (Boccaletti et al., 2006; Newman et al., 2006; Newman, 2010). These characteristics are based on the idea which I would like to call the *real view* on networks: each node is just a point and edges or arcs between nodes indicate the existence of some kind of interaction between nodes if a system is represented as a network. However, in many real systems, it is the case that some kind of process is running within an object represented by a node. For example, in neuronal networks, nodes are neurons that have information processing ability. In gene regulation networks, nodes are genes, but we should include proteins coded by those genes into nodes if we consider regulation relationships as arcs. Thus, we can think that complicated chemical processes to synthesize proteins occur within each node in a gene regulation network. We can interpret other biological networks including ecological networks, metabolic networks in the same way. If we consider objects as processes, then interactions between objects can be seen as interfaces between processes. I would like to call this view “objects as processes, interactions as interfaces” on networks the *dual view* in contrast to the ordinary real view mentioned above.

In this paper, we examine what is involved in having internal processes on nodes in general for complex networks. Usually, processes occurring on nodes are described as particular dynamics. Then, an appropriate statistical ensemble of dynamics is studied in order to conclude something in general (e.g. random Boolean networks by (Kauffman, 1969)). Instead of statistical generality, we here appeal to category theoretical universality to study the problem.

We note that there is an inverse dual view, namely, “processes as objects”. This idea appears in the formulation of Metabolism-Repair System by R. Rosen (Rosen, 1958). Recently, the idea was used as the line graph transformation in the community detection problem in complex networks (Ahn et al., 2010; Evans and Lambiotte, 2009). The two ideas “objects as processes” and “processes as objects” have a certain dual relationship called category theoretical adjunction (MacLane, 1998) if they are formalized within the framework of category theory (Haruna and Gunji, 2007; Pultr, 1979).

There are many ways (indeed, uncountably many ways) to consider objects as processes. However, we can show that there exists a canonical way (in a precise mathematical sense stated in Section 3) among all the ways to see objects as processes within the framework of category theory (Haruna, 2011b). The canonical way to see objects as processes gives rise to an equivalence relation on the set of arcs of each directed network. This equivalence relation can be interpreted as defining a new notion of connectedness called *lateral con-*

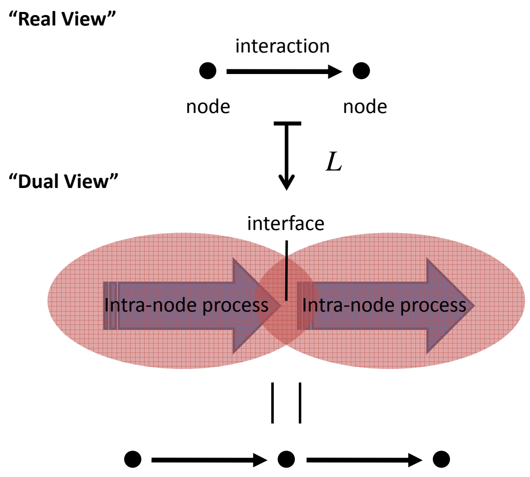

Figure 1: The idea "objects as processes, interactions as interfaces".

nectedness. An intuitive explanation of derivation of lateral connectedness without category theory is the main aim of the former half of this paper. In the latter half, we analyze the neuronal network of the nematode *Caenorhabditis elegans* based on lateral connectedness as a first application to real world networks.

This paper is organized as follows. In section 2, we describe a mathematical formulation of the dual view on directed networks. In section 3, we introduce lateral connectedness for directed networks as a naturally emerging structure from the dual view. In section 4, we apply lateral connectedness to the neuronal network of *C. elegans* and discuss its functional significance. In section 5, we give conclusions and outlooks.

Objects as Processes, Interactions as Interfaces

In this paper, we only consider directed networks. Some early attempts related to the content of this section are found in Haruna and Gunji (2007); Haruna (2008a,b, 2011a).

In the dual view introduced in Section 1, each node is interpreted as a process and each arc is seen as an interface between two processes. This idea can be formalized as network transformations as follows.

As a motivating example, let us interpret each node as an arc (together with its source and target nodes) representing a process running in the node and each arc as a node connecting two arcs representing processes running in the original two nodes (Fig. 1). Of course, each node can be replaced by a much more complicated network representing a process running within the node. The connection between the two complicated networks can also be arbitrary. We call a network (that can be arbitrary complicated) representing a process running within a node together with information how its two copies form an interface corresponding to an arc a *model of directed network type*. In general, models of directed network type need not consist of directed networks (Haruna, 2011b), however, in the following discussion, we restrict ourselves on models consisting of directed networks for simplicity.

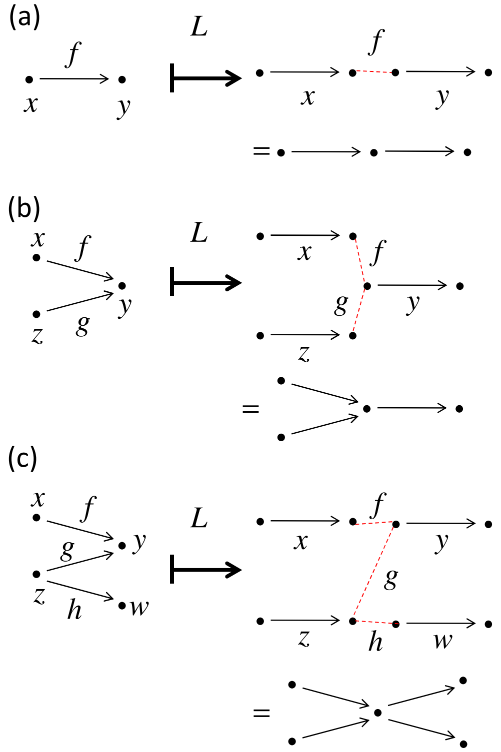

Figure 2: Three examples of the calculation of the network transformation L.

Fig. 2 illustrates how the above motivating model of directed network type gives rise to a network transformation L. In Fig. 2 (a), the two nodes x and y are converted to two arcs x and y by L. The target of x and the source of y are glued by the arc f in the original network. In Fig. 2 (b), there are three copies of arcs x, y and z after the transformation L corresponding to the three nodes x, y and z in the original network. Their sources and targets are glued according to the arcs f and g in the original network. The similar copy and glue rule works for the example in Fig. 2 (c).

Formally, the network transformation L can be defined as follows. Let $G = (A, O, \partial_0, \partial_1)$ be a directed network, where A is a set of arcs, O is a set of nodes and ∂_0 and ∂_1 are maps from the set A to the set O that send each arc to its source node and target node, respectively. The directed network $L(G)$ obtained by the application of L to G is a quartet $L(G) = (O, O \times \{0,1\}/\sim, \partial'_0, \partial'_1)$, where the set of arcs of $L(G)$ is identical to the set of nodes O of G, the set of nodes of $L(G)$ is a quotient set $O \times \{0,1\}/\sim$ and

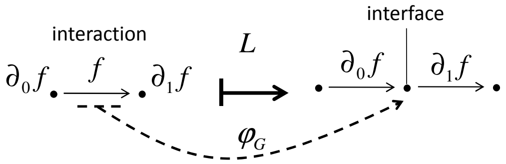

Figure 3: The map φ_G materialize the idea "interaction as interface".

$\sim$ is an equivalence relation on the set $O \times \{0,1\}$ generated by the relation defined by $(x,1) \sim (y,0)$ if and only if there exists an arc f from x to y in G. The symbol 1 indicates the "source part" of the node x and the symbol 0 indicates the "target part" of the node y. The source and target maps ∂_0', ∂_1' are defined naturally.

In general, for any model of directed network type, the induced network transformation can be described by a similar copy and glue rule, no matter how complicated it is. For a category theoretical formulation, see (Haruna, 2011a).

A New Notion of Connectedness

By the network transformation L introduced in Section 2, each node in a directed network G is sent to an arc in $L(G)$. On the other hand, we can think that each arc f in a directed network $G = (A, O, \partial_0, \partial_1)$ is mapped to a node in $L(G)$ between two arcs in $L(G)$ corresponding to the source and the target nodes of f, namely, $\partial_0 f$ and $\partial_1 f$ (Fig. 3). We denote this map by $\varphi_G : A \to O \times \{0,1\}/\sim$. For each arc $f \in A$, $\varphi_G(f)$ is the target of $\partial_0 f$ (or the source of $\partial_1 f$) in $L(G)$. Hence, we have $\varphi_G(f) = [(\partial_0 f, 1)](= [(\partial_1 f, 0)])$, where $[(x,i)]$ is an equivalence class containing $(x,i) \in O \times \{0,1\}$.

A natural question about the nature of the map φ_G is "When does $\varphi_G(f) = \varphi_G(g)$ hold for arcs $f, g \in A$?" The answer is straightforward and the necessary and sufficient condition for the equality $\varphi_G(f) = \varphi_G(g)$ is that there exists a zigzag sequence of arcs between f and g as indicated in Fig. 4. We say that two arcs f and g are *laterally connected* if $\varphi_G(f) = \varphi_G(g)$ holds.

For any model of directed network type, a similar map on the set of arcs of a given directed network can be defined. Such a map induces an equivalence relation on the set of arcs by identifying two arcs if they are sent to the same element in the codomain of the map. Let us denote the equivalence relation induced by the map φ_G above by R_{LC}. Then, R_{LC} is canonical in the following sense. *For any directed network $G = (A, O, \partial_0, \partial_1)$, R_{LC} is the smallest equivalence relation on the set of arcs A among those induced by all models of directed network type.* In other words, the partition of the set of arcs induced by lateral connectedness is the finest one among those induced by the idea "objects as processes, interactions as interfaces". We call each equivalence class *laterally connected component*. This statement can be proved within the framework of category theory in more general form (Haruna, 2011b).

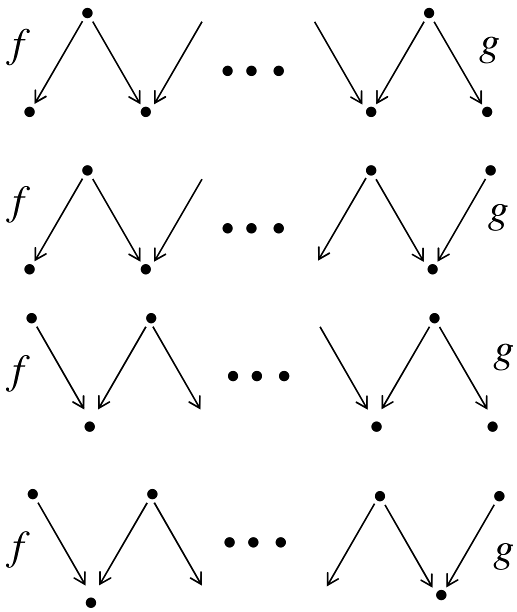

Figure 4: Two arcs f and g are laterally connected if there is a zigzag sequence of arcs between them. There are four cases depending on the situations at the both ends.

In summary, we obtain the notion of lateral connectedness as a canonically emerging structure of directed networks from the idea "objects as processes, interactions as interfaces".

By its definition, lateral connectedness may be relevant with functional commonality between arcs emerging from shared input or output. This is in contrast to the notion of strong connectedness. Here we say that two arcs are strongly connected if one arc can be reached from the other by a directed path and vice versa. Strong connectedness may be associated with functionality resulting from circulation of information or materials. Intuitively, they seem to be dual to each other. Indeed, this intuition can be enhanced by the following category theoretical point of view.

Lateral connectedness derives from the network transformation L which is based on the idea "objects as processes, interactions as interfaces". On the other hand, strong connectedness can be obtained from the line graph transformation R which is based on the idea "processes as objects". Given a directed network $G = (A, O, \partial_0, \partial_1)$, its line graph is a directed network $R(G) = (S, A, \partial_0'', \partial_1'')$, where $S = \{(f,g) \in A \times A | \partial_1 f = \partial_0 g\}$, $\partial_0''(f,g) = f$ and $\partial_1''(f,g) = g$ for any $(f,g) \in S$. As noted in Section 1, the two transformations L and R satisfy a certain category theoretical duality called adjunction (Haruna and Gunji, 2007;

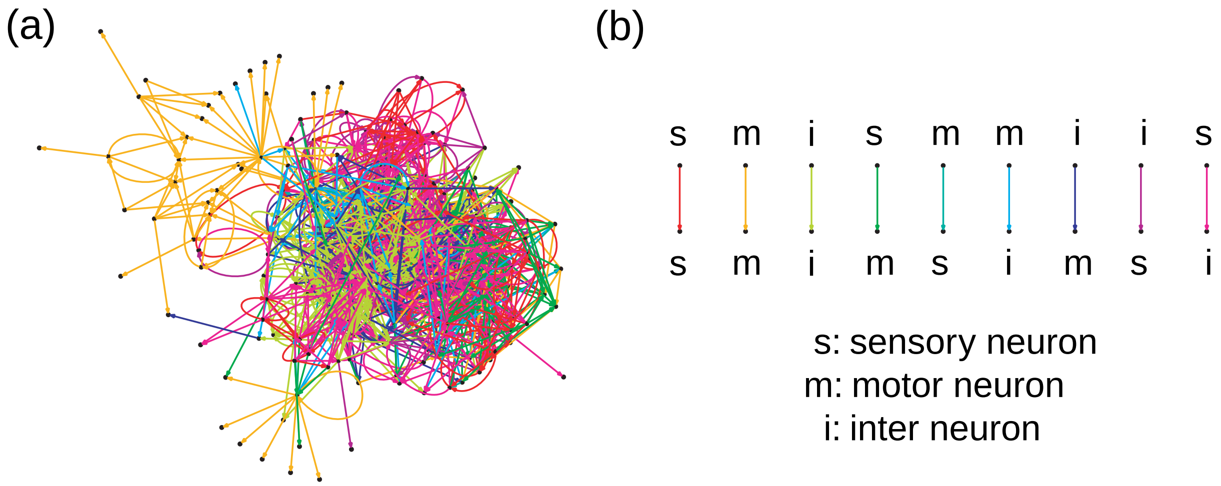

Figure 5: The wiring diagram of the neuronal network of *C. elegans* based on the database (Oshio et al., 2003) (http://ims.dse.ibaraki.ac.jp/ccep/) depicted by Graphviz (http://www.graphviz.org/). (a) Arcs are colored based on pairs of functions of their source and target neurons. (b) Correspondence between colors and pairs of functions of neurons.

Pultr, 1979). By definition, S is the set of arcs of the directed network $R(G)$, but it can be seen as a binary relation on the set A. Mathematically, the notion of strong connectedness defined above is an equivalence relation SC on the set A of arcs of G. On the other hand, we have an equality

$$SC = \overline{S} \cap \overline{S^{-1}}, \tag{1}$$

where S^{-1} is the inverse of the binary relation S and $\overline{T}$ for a binary relation T on A is its reflexive and transitive closure. In this sense, strong connectedness is generated by the line graph transformation R which is category theoretical dual to L.

One might think that the duality between lateral connectedness and strong connectedness in the above sense is a mathematical expression for Lorente de Nó's two principles of plurality and reciprocity (Lorente de Nó, 1938).

Analysis of a Neuronal Network

In this section, we discuss significance of lateral connectedness in the neuronal network of *C. elegans* as a first application of it. We compare the partitions of the set of arcs based on functions of neurons with the partition induced by lateral connectedness to examine functional significance of lateral connectedness. We make use of two similarity measures described in the next subsection for the comparison.

Network Data

We make use of the database constructed by Oshio et al. (2003) (http://ims.dse.ibaraki.ac.jp/ccep/) whose original reference is White et al. (1986). We remove nodes and connections other than neurons and chemical synapses. The resulting data set contains 233 neurons among 282 neurons in the somatic nervous system and 4170 chemical synapses. We construct a family of directed networks whose nodes are 233 neurons in the following way: First, we put an arc from one node to another node if there exists a chemical synapse from the former to the latter. Second, since there is multiple chemical synapses from one neuron to another neuron in general, we specify a weight for each arc by the number of chemical synapses from the source to the target of the arc. Finally, we introduce thresholds for the weight values and consider the network topology consisting of arcs whose weights are greater or equal to a given threshold.

Each neuron has one of three functional types: sensory, inter and motor. We consider three partitions of the set of arcs based on the functions of neurons. The first one is called *ST-partition* which considers types of the two neurons at both ends of each arc. Thus, there are nine clusters for the ST-partition. In the wiring diagram shown in Fig. 5 (a) where threshold is 1, each arc is colored based on the ST-partition. The correspondence between colors and the ST-partition clusters is indicated in Fig. 5 (b). The second one is called the *S-partition* which considers type of the source neuron of each arc. The third one is called the *T-partition* which considers type of the target neuron of each arc. The number of clusters in the S-partition or T-partition is three.

The equivalence relation R_{LC} induced by lateral connectedness also gives rise to a partition of the set of arcs. We call this partition the *LC-partition*. In the following discussion, we measure similarity between the LC-partition and the above three functional partitions.

Similarity Measures

We make use of two similarity measures to quantify similarity between two partitions on a set. The first one is called the

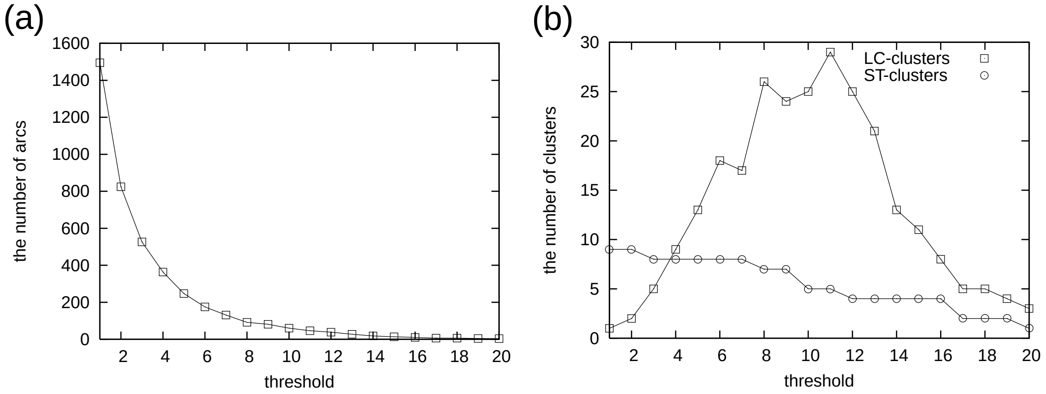

Figure 6: (a) The number of arcs as a function of threshold. (b) The number of clusters in the LC-partition and in the ST-partition as a function of threshold.

Adjusted Rand Index (ARI) (Hubert and Arabie, 1985). The second one is called the *Adjusted normalized Mutual Information (AMI)* (Vinh et al., 2009). To explain the idea of the ARI, we first review the definition of the *Rand Index (RI)* (Rand, 1971).

Let X be a set consisting of N points. Let $\mathbf{U} = \{U_1, U_2, \cdots, U_l\}$ and $\mathbf{V} = \{V_1, V_2, \cdots, V_m\}$ be two partitions of X, namely, they are families of subsets of X satisfying $U_i \cap U_{i'} = V_j \cap V_{j'} = \emptyset$ for $i \neq i'$, $j \neq j'$ and $\cup_{i=1}^{l} U_i = \cup_{j=1}^{m} V_j = X$. Let us put $n_{ij} := |U_i \cap V_j|$, $a_i := |U_i|$ and $b_j := |V_j|$ for $i = 1, 2, \cdots, l$ and $j = 1, 2, \cdots, m$, where $|Y|$ for a set Y denotes its cardinality. Then, we have $a_i = \sum_{j'=1}^{m} n_{ij'}$ and $b_j = \sum_{i'=1}^{l} n_{i'j}$ for $i = 1, 2, \cdots, l$ and $j = 1, 2, \cdots, m$. A $l \times m$ matrix $C := (n_{ij})$ is called the *contingency matrix*, which encodes information how two partitions $\mathbf{U}$ and $\mathbf{V}$ overlap. We can calculate both the ARI and the AMI by using elements of the contingency matrix C.

The Rand Index (RI) between partitions $\mathbf{U}$ and $\mathbf{V}$ is defined by counting the number of pairs of elements of X on which two partitions agree or disagree:

$$RI(\mathbf{U}, \mathbf{V}) = \frac{N_{00} + N_{11}}{N_{00} + N_{01} + N_{10} + N_{11}}, \tag{2}$$

where N_{00} is the number of pairs that are in the same cluster in both $\mathbf{U}$ and $\mathbf{V}$, N_{01} is the number of pairs that are in the same cluster in $\mathbf{U}$ but in different clusters in $\mathbf{V}$, N_{10} is the number of pairs that are in different clusters in $\mathbf{U}$ but in the same cluster in $\mathbf{V}$ and N_{11} is the number of pairs that are in different clusters in both $\mathbf{U}$ and $\mathbf{V}$. After a few algebras, one can see that N_{01} and N_{10} are given by

$$N_{01} = \sum_{j=1}^{m} \binom{b_j}{2} - \sum_{i=1}^{l} \sum_{j=1}^{m} \binom{n_{ij}}{2}, \tag{3}$$

$$N_{10} = \sum_{i=1}^{l} \binom{a_i}{2} - \sum_{i=1}^{l} \sum_{j=1}^{m} \binom{n_{ij}}{2}. \tag{4}$$

Since we have $N_{00} + N_{01} + N_{10} + N_{11} = \binom{N}{2}$, we obtain the following explicit formula for the RI:

$$RI(\mathbf{U}, \mathbf{V}) = \frac{\binom{N}{2} - \left\{ \sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right\} + 2 \sum_{i,j} \binom{n_{ij}}{2}}{\binom{N}{2}}. \tag{5}$$

The RI takes its maximum value 1 when two partitions are identical. The minimum value 0 is taken if and only if one partition consists of a single cluster and the other consists of only clusters with a single point, which is hard to satisfy by random partitions. Indeed, the RI takes relatively high values for two random partitions. However, it is plausible for a similarity measure to take values close to zero for random partitions. To improve this disadvantage of the RI, Hubert and Arabie (1985) introduced the Adjusted Rand Index (ARI) which takes over a correction for chance:

$$ARI(\mathbf{U}, \mathbf{V}) = \frac{RI(\mathbf{U}, \mathbf{V}) - E\left(RI|\mathbf{a}, \mathbf{b}\right)}{1 - E\left(RI|\mathbf{a}, \mathbf{b}\right)}, \tag{6}$$

where 1 in the denominator is the maximum value of the RI and $E\left(RI|\mathbf{a}, \mathbf{b}\right)$ is the expected value of the RI between two randomly chosen partitions of the set X subject to the condition that two vectors $\mathbf{a} = (a_1, a_2, \cdots, a_l)$ and $\mathbf{b} = (b_1, b_2, \cdots, b_m)$ are fixed. Since we have $E\left(\binom{n_{ij}}{2}|\mathbf{a}, \mathbf{b}\right) = \binom{a_i}{2}\binom{b_j}{2}/\binom{N}{2}$ (Hubert and Arabie, 1985), an explicit formula for the ARI is given by

$$ARI(\mathbf{U}, \mathbf{V}) = \frac{\sum_{i,j} \binom{n_{ij}}{2} - \left\{ \sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right\} / \binom{N}{2}}{\frac{1}{2} \left\{ \sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} \right\} - \left\{ \sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2} \right\} / \binom{N}{2}}. \tag{7}$$

Our second measure of similarity, the AMI is defined based on the mutual information between two partitions (Vinh et al., 2009). Let us introduce the probability that an element of X is contained in a cluster U_i by $P(i) = a_i/N$.

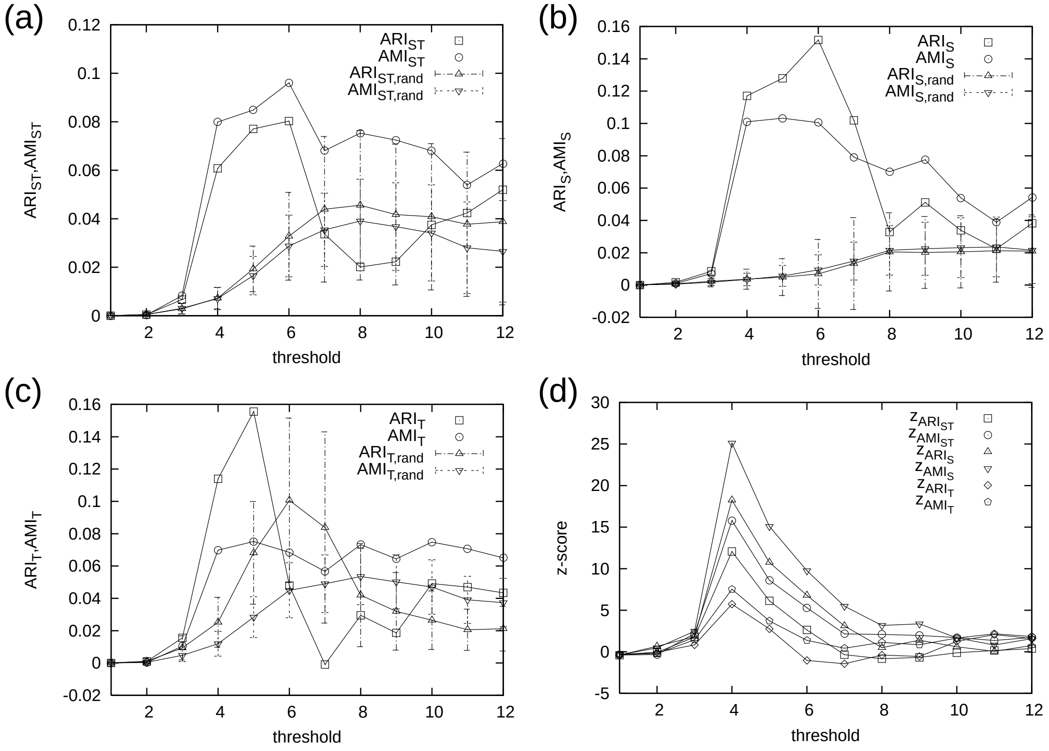

Figure 7: Comparisons between the LC-partition and partitions based on neuron functions. Two similarity measures, one is pair-counting based (the ARI) and the other is information-theoretic (the AMI), are used. (a) The LC-partition vs the ST-partition. (b) The LC-partition vs the S-partition. (c) The LC-partition vs the T-partition. (d) Z-scores for the values of two similarity measures as functions of threshold.

The Shannon entropy with respect to the partition $\mathbf{U}$ is defined by $H(\mathbf{U}) = -\sum_{i=1}^{l} P(i)\log_2 P(i)$. Similarly, the Shannon entropy with respect to the partition $\mathbf{V}$ is given by $H(\mathbf{V}) = -\sum_{j=1}^{m} P'(j)\log_2 P'(j)$, where $P'(j) = b_j/N$. Then, the mutual information between two partitions $\mathbf{U}$ and $\mathbf{V}$ is defined by $I(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{l}\sum_{j=1}^{m} P(i,j)\log_2 \frac{P(i,j)}{P(i)P'(j)}$, where $P(i,j) = n_{ij}/N$ which is the joint probability that an element of X falls into both U_i and V_j.

Strehl and Ghosh (2002) proposed the normalized mutual information (NMI) as follows:

$$NMI(\mathbf{U},\mathbf{V}) = \frac{I(\mathbf{U},\mathbf{V})}{\sqrt{H(\mathbf{U})H(\mathbf{V})}}, \tag{8}$$

which takes its values in the unit interval $[0,1]$. The NMI takes its maximum value 1 when two partitions are identical. The minimum value 0 is realized when two partitions are independent, namely, $n_{ij} = a_i b_j$ holds for all $1 \le i \le l$ and $1 \le j \le m$. Hence, the NMI for random partitions takes its values close to 0. However, its adjusted version is more preferable. The adjusted normalized mutual information (AMI) is defined in the similar spirit as in the ARI (Vinh et al., 2009) :

$$AMI(\mathbf{U},\mathbf{V}) = \frac{I(\mathbf{U},\mathbf{V}) - E\left(I|\mathbf{a},\mathbf{b}\right)}{\sqrt{H(\mathbf{U})H(\mathbf{V})} - E\left(I|\mathbf{a},\mathbf{b}\right)}, \tag{9}$$

where $E\left(I|\mathbf{a},\mathbf{b}\right)$ is the expected value of the mutual information I between two randomly chosen partitions of the set X subject to the condition that two vectors $\mathbf{a} = (a_1, a_2, \cdots, a_l)$ and $\mathbf{b} = (b_1, b_2, \cdots, b_m)$ are fixed.

In the next subsection, we apply these two adjusted similarity measures, the ARI and the AMI, to the partitions of the set of arcs in the neuronal network of *C. elegans* by neuron functions and the partition based on lateral connectedness for each threshold.

Results

Fig. 6 (a) shows the number of arcs as a function of threshold. Fig. 6 (b) indicates the number of clusters in the LC-partition and in the ST-partition. The former tends to increase for thresholds within the range from 1 to 10 because decrease in the number of arcs can leads to division of one cluster into two or more clusters. It decreases for thresholds larger than 12 simply because the number of arcs is too small

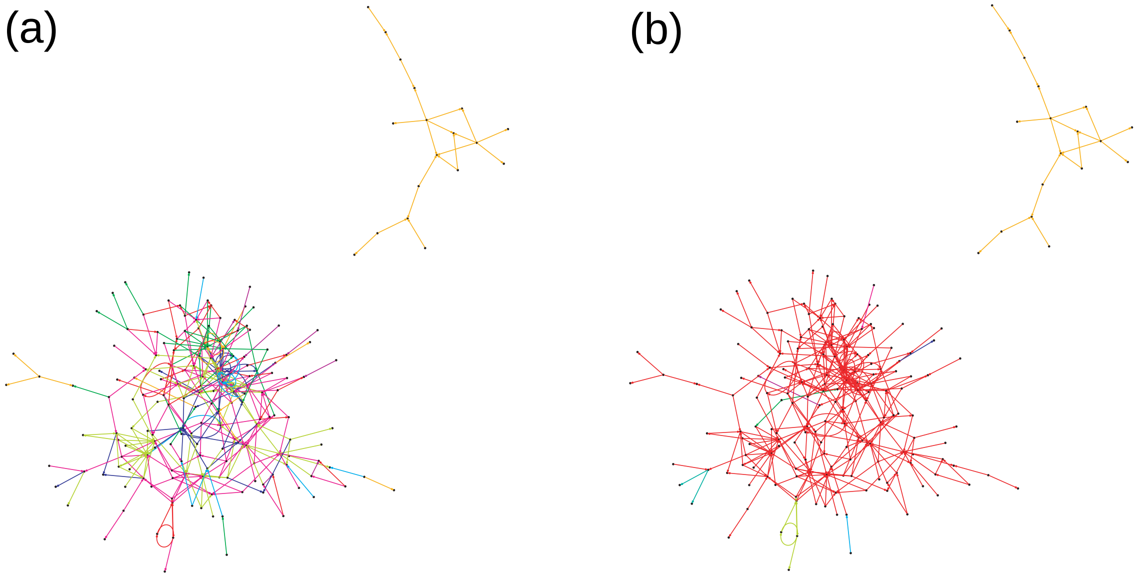

Figure 8: The wiring diagram of the neuronal network of *C. elegans* at threshold 4 depicted by Graphviz (http://www.graphviz.org/). (a) Arcs are colored by the ST-partition. Color assignment is the same for Fig. 5. (b) Arcs are colored by the LC-partition. Colors of different clusters are specified arbitrarily.

for the number of the LC-clusters to grow by divisions for those relatively large thresholds.

In Fig. 7 (a), we plot the ARI and the AMI between the LC-partition and the ST-partition as a function of threshold. It takes its maximum value when threshold is equal to 6. As a control experiment, we calculate averages and standard deviations of the ARI and the AMI between the LC-partition and the ST-partition on 1000 randomized networks by re-wiring arcs randomly, which are also shown in Fig. 7 (a). Note that degree distributions are invariant under the re-wiring process. We can see a large deviation from the control around the maximum point. We have similar results for the S-partition and the T-partition (Fig. 7 (b),(c)).

To quantify deviation from the control experiment, we calculate the Z-score for each comparison. The Z-score of a quantity Q is defined by

$$z_Q = \frac{Q_{orig} - \langle Q_{rand} \rangle}{\sigma}, \tag{10}$$

where Q_{orig} is the value of Q in the original network, $\langle Q_{rand} \rangle$ is the average of Q calculated from an ensemble of randomized networks and σ is its standard deviation. The Z-scores of both the ARI and the AMI take their maximum value when threshold is equal to 4 for all comparisons (Fig. 7 (d)). All of the maximum values of the Z-scores are more than 5, which indicates significant deviation from the control in all comparisons. However, we should note that the absolute values of the two similarity measures are not so high, at most 0.152.

In Fig. 8, we plot the wiring diagram where we set threshold 4. Arcs are colored based on the ST-partition (Fig. 8 (a)) and the LC-partition (Fig. 8 (b)). There are two weakly connected components, one is large and the other is small. Here, we define a weakly connected component of a directed network as a maximal set of arcs in which every pair of arcs are connected by a sequence of arcs ignoring the direction. Almost all the motor-motor connections are included in the smaller weakly connected component on one hand, they also form a single laterally connected component in the LC-partition. However, the LC-partition fails to capture more detailed functional partition within the larger weakly connected component possibly due to many recurrent connections between the sensor, inter and motor region of the neuronal network of *C. elegans* (Varshney et al., 2011). This is one reason that we have relatively low absolute values for the two similarity measures.

Conclusions and Outlooks

In this paper, we intuitively explained that how the idea "objects as processes, interactions as interfaces" can be formalized within the framework of category theory. We derived the notion of lateral connectedness as a canonical structure obtained from the idea. By its definition, lateral connectedness has possibility to be associated with functional commonality between arcs arising from shared input or output. As a first application of lateral connectedness, we examined functional significance of lateral connectedness in the neuronal network of *C. elegans* by the method of clustering comparison. For the analysis, we made use of two similarity measures to quantify similarity between two partitions on the same set, one is pair-counting based measure and the other is information-theoretic measure.

We showed that the partition of the set of arcs based on lateral connectedness is not inconsistent with the functional partition of the set of arcs. However, even if we set threshold at the point where the largest deviation from an ensemble of randomized networks is observed, it can only capture a part of the partitions based on neuron functions. One problem of the analysis performed in this paper may be that the direct comparison to functional partitions is too strict to recognize significance of lateral connectedness. Another problem is that the data used is incomplete. Analysis with more complete data (Varshney et al., 2011) will be necessary.

Introduction of lateral connectedness has several implications. First, we can analytically solve percolation problems with respect to lateral connectedness on configuration model (networks chosen uniformly at random from the set of all possible networks with a specified degree distribution) of directed networks (Haruna, 2011b). Applications of the analytical result on configuration model to biological networks are now ongoing. Second, we can define alternatives for some notions used in conventional complex network studies. For example, the notion of path length can be defined based on lateral connectedness. Since metrics such as closeness and betweenness centralities are functions of path lengths, they are also the targets of alternative definitions. Finally, theoretical development and empirical applications of the duality between lateral connectedness and strong connectedness are also intriguing issues.

Acknowledgements

This work was supported by JST PRESTO program.

References

Ahn, Y.-Y., Bagrow, J. P., and Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. *Nature*, 466:761–765.

Alon, U. (2006). *Inroduction to Systems Biology: Design Principles of Biological Circuits*. CRC Press, Boca Raton.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006). Complex networks: Strucutre and dynamics. *Phys. Rep.*, 424:175–308.

Evans, T. S. and Lambiotte, R. (2009). Line graphs, link partitions, and overlapping communities. *Phys. Rev. E*, 80:016105.

Haruna, T. (2008a). Algebraic theory of biological organization. Doctral Dissertation, Kobe University.

Haruna, T. (2008b). Being arranged in advance: Quantum entanglement and biological feedback. In Bullock, S., Noble, J., Watson, R., and Bedau, M. A., editors, *Artificial Life XI: Proceedigns of th Eleventh International Conference on the Simulation and Synthesis of Living Systems*, pages 220–226. MIT Press, Cambridge, MA.

Haruna, T. (2011a). An application of category theory to the study of complex networks. Int. J. Comp. Anti. Sys. (in press).

Haruna, T. (2011b). In preparation.

Haruna, T. and Gunji, Y.-P. (2007). Duality between decomposition and gluing: A theoretical biology via adjoint functors. *BioSystems*, 90:716–727.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *J. Classification*, 2:193–218.

Junker, B. H. and Schreiber, F. (2008). *Analysis of Biological Networks*. John Wiley & Sons, Inc.

Kauffman, S. A. (1969). Metablic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.*, 22:437–467.

Lorente de Nó, R. (1938). Analysis of the activity of the chains of internuncial neurons. *J. Neurophysiol.*, 1:207–244.

MacLane, S. (1998). *Categories for the Working Mathematician, 2nd edition*. Springer-Verlag, New York.

Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford Univ. Press Inc., New York.

Newman, M. E. J., Barabási, A.-L., and Watts, D. J. (2006). *The Structure and Dynamics of Networks*. Princeton Univ. Press, Princeton.

Oshio, K., Iwasaki, Y., Morita, S., Osana, Y., Gomi, S., Akiyama, E., Omata, K., Oka, K., and Kawamura, K. (2003). Database of synaptic connectivity of c. elegans for computation. Technical Report of CCeP, Keio Future, No. 3, Keio University.

Pultr, A. (1979). On linear representations of graphs. In *Fundamentals of computation theory(Proc. Conf. Algebraic, Arith. and Categorical Methods in Comput. Theory, Berlin/Wendisch-Riets, 1979), Math. Res. 2*, pages 362–369. Akademie-Verlag, Berlin.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. Am. Stat. Assoc.*, 66:846–850.

Rosen, R. (1958). The representation of biological systems from the standpoint of the theory of categories. *Bull. Math. Biophys.*, 20:317–341.

Sporns, O. (2011). *Networks of the Brain*. MIT Press, Cambridge, MA.

Strehl, A. and Ghosh, J. (2002). Cluster ensembles - a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.*, 3:583–617.

Varshney, L. R., Chen, B. L. Paniagua, E., Hall, D. H., and Chklovskii, D. B. (2011). Structural properties of the caenorhabditis elegans neuronal network. *PLoS computational biology*, 7:e1001066.

Vinh, N. X., Epps, J., and Bailey, J. (2009). Information theoretic measures for clusterings comparison: is a correction for chance necessary? In Danyluk, A., Bottou, L., and Littman, M., editors, *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1073–1080. ACM New York, NY, USA.

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode caenorhabditis elegans. *Phil. Trans. R. Soc. London B*, 314:1–340.