Decomposability of Multivariate Interactions

Naoki Miyagawa[1,2]
Hiroshi Teramoto[1,3]
Chun-Biu Li[1,2,4]
Tamiki Komatsuzaki[1,3,4]

[1]*Research Institute for Electronic Science*
Hokkaido University, Sapporo 001-0020, Japan

[2]*Department of Mathematics*
Hokkaido University, Sapporo 060-0810, Japan

[3]*Graduate School of Life Science*
Hokkaido University, Sapporo 060-0810, Japan

[4]*Research Center for Integrative Mathematics*
Hokkaido University, Sapporo 060-0812, Japan
g05041037@gmail.com, teramoto@es.hokudai.ac.jp, cbli@es.hokudai.ac.jp,
tamiki@es.hokudai.ac.jp

Systems in nature, composed of many microscopic components, exhibit several distinctive global patterns. Can we understand the emergent patterns in terms of the components? One possible device to address such a question is to scrutinize the "interactions" among these components, from which the global behavior arises. In this paper, we introduce and generalize the information-theoretic quantity called connected information. It provides us with a measure of many-body interactions buried in complex systems. While the original connected information is formulated globally to include all contributions from the microscopic components, we formulate decomposition rules for the connected information to capture local interactions. The implication of our results will also be discussed in relation to the identification of local functional modules in neural systems based on experimental observations.

1. Introduction

In statistical mechanics, the probability distributions originated from quantum mechanics and/or chaotic dynamics provide us with the fundamental description in connecting the microscopic dynamics (or the first principles) and the macroscopic dynamics (or thermodynamics). In order to understand the dynamical hierarchies that exist in nature, statistical mechanics requires that the first principles' equation of motions of the composite elements are given in the system of interest. However, this is in general not the case, especially for systems in biological and social sciences. This raises the question: how can we discuss the connection between different dynamical hierarchies for those

systems whose microscopic dynamical rules are not known? One way to resolve this problem is to extract the underlying microscopic rules by examining the properties of multivariate dependence in the observed probability distributions of the composite elements. In particular, Schneidman et al. recently introduced a quantity called "connected information" that quantifies the strength of many-body interactions, such as pairwise, triple, quadruple, and so on, among composite elements in the given system [1].

The connected information was originally formulated to capture only global spatial properties, namely, the k-body interactions containing contributions from the set of all possible combinations in k bodies in the system. In this paper, we formulate decomposition rules to express the global interactions as a superposition of local interactions that cannot be decomposed further. These indecomposable local interactions can then serve as the fundamental building blocks of the multivariate systems. In Section 2, we will first review connected information constructed from the multivariate probability distributions. Some mathematical properties of the connected information and the local decomposition of the global interactions will be given in Section 3, followed by the Conclusion in Section 4.

2. Connected Information and the Multivariate Dependence

We first describe the connected information that measures the strength of correlation of different orders among the multiple variables in the system. Let us consider N random discrete variables $x = (x_1, \ldots, x_N)$ whose joint probability distribution is denoted by $P(x)\,(= P(x_1, \ldots, x_N))$. For simplicity, we consider that x are binary variables, that is, $x_j \in \{0, 1\}$. The generalization to nonbinary cases is straightforward. In terms of the joint probability, the Shannon entropy $H[P]$ is defined by

$$H[P] := -\sum_x P(x) \log P(x), \tag{1}$$

to measure the information content that is missing in order to predict the value of x [2]. Here $\sum_x$ denotes the summation over all variables in x. On the other hand, mutual information is a well-known measure of statistical dependency among the variables given by

$$I_m[P] := \sum_x P(x) \log \left[\frac{P(x)}{\prod_j P_j(x_j)} \right] = \sum_{j=1}^N H[P_j] - H[P], \tag{2}$$

where

$$P_j(x_j) = \sum_{x_1} \cdots \overset{\wedge}{\sum_{x_j}} \cdots \sum_{x_N} P(x_1, \ldots, x_N). \tag{3}$$

The notation $\hat{\Sigma}_{x_j}$ means that the summation over x_j is excluded and $P_j(x_j)$ $(j = 1, \ldots, N)$ are referred to as the first order marginals. $H[P_j]$ is the entropy of the single variable x_j. Since mutual information only measures the difference between the information content of the full joint probability $P(x_1, \ldots, x_N)$ and that of $\prod_j P_j(x_j)$ corresponding to the fully statistically independent case, it cannot tell us how much of the mutual information can be explained by pairwise, triple, and higher-order correlations (interactions).

To quantify, for example, the pairwise interaction among the variables of a system similar to equation (2), consider a quantity $\tilde{I}^{(2)}$ defined by

$$\tilde{I}^{(2)}[P] := \sum_i H[P_i] - \frac{1}{N-1} \sum_{i<j} H[P_{ij}], \tag{4}$$

where P_{ij} are called the second order marginals given by

$$P_{ij}(x_i, x_j) := \sum_{x_1} \cdots \hat{\sum_{x_i}} \cdots \hat{\sum_{x_j}} \cdots \sum_{x_N} P(x_1, \ldots, x_N). \tag{5}$$

However, $\tilde{I}^{(2)}[P]$ may not vanish when the joint probability distribution $P(\boldsymbol{x})$ has no pairwise interactions. Here, k-body interactions such as pairwise, triple, and quadruple mean that the probability distribution $P(\boldsymbol{x})$ can be decomposed into a product of functions of only k variables (e.g., x_i and x_j for pairwise interaction), which cannot be further divided into functions of fewer variables. Thus, no pairwise interaction in the system implies that either $P(\boldsymbol{x})$ is written as a product of one-variable functions $P_i(x_i)$, that is, all variables are independent of each other, or $P(\boldsymbol{x})$ is written as a product of functions of more than two variables that cannot be reduced into functions of fewer variables. Note that in the latter case $\tilde{I}^{(2)}$ can still have a nonzero value while it has zero in the former case. Therefore, we introduce the following probability distribution

$$P^{(k)}(\boldsymbol{x}) := \underset{Q \in M_k}{\arg\max} H[Q], \tag{6}$$

which we call the kth order maximum entropy distribution (MED). This distribution maximizes the entropy on a space M_k defined by

$$M_k = \Big\{ Q \Big| \sum_{x_{i_1}, \ldots, x_{i_{N-k}}} Q(\boldsymbol{x}) = \sum_{x_{i_1}, \ldots, x_{i_{N-k}}} P(\boldsymbol{x}),$$

$$\forall\, i_1 < \cdots < i_{N-k} \in \{1, \ldots, N\} \Big\}. \tag{7}$$

This space is called the kth marginal equivalence space in which all distributions $Q(\boldsymbol{x})$ have the same k^{th} order marginals as the original $P(\boldsymbol{x})$. Importantly, maximization of the entropy in M_k means that the kth order MED (equation (6)) carries solely the part of information of the joint probability distribution $P(\boldsymbol{x})$ associated with the kth order marginals. However, the information of the kth order marginals contains that of the $(k-1)^{\text{th}}$ order marginals because $M^{k-1} \subset M^k$. In order to capture solely the information associated with exactly k variables but not with fewer, it is necessary to subtract the contribution of $(k-1)$ variables from the information of k variables, resulting in a quantity termed the kth order connected information $I_C^{(k)}[P]$:

$$I_C^{(k)}[P] := \left(-H\left[P^{(k)}\right]\right) - \left(-H\left[P^{(k-1)}\right]\right). \tag{8}$$

$I_C^{(k)}[P]$ vanishes if and only if the probability distribution $P(\boldsymbol{x})$ has no k-body interactions.

It can be easily verified that equation (2) can be hierarchically decomposed into the connected information of different orders as

$$I_m[P] = \sum_{k=2}^{N} I_C^{(k)}[P]. \tag{9}$$

Since the connected information of order k quantifies the k-body interactions among all N variables in the system, equation (9) corresponds to the "global" decomposition that cannot further resolve the "local" contribution to $I_C^{(k)}[P]$ from a particular set of k variables. For example, the second order connected information $I_C^{(2)}[P]$ for a system of three variables consists of all pairwise interactions associated with the second order marginals (i.e., $P_{12}(x_1, x_2)$, $P_{23}(x_2, x_3)$, and $P_{31}(x_3, x_1)$) that cannot allow us to evaluate how much contribution(s) arise(s) from a specific individual or set of pairs of variables (e.g., from the pair x_1 and x_3 only). Therefore, it is natural to ask if the concept of connection information can be generalized to identify local interactions of a given order. This generalization becomes crucial and necessary when the system possesses strong spatial heterogeneity. In the following, we will scrutinize the relationship between global information that contains all pairwise interactions $I_C^{(2)}[P]$ and local information that contains solely the individual pairwise interactions. It was found that interactions among neurons in neural networks can be well described by pairwise interactions even when the neuronal connectivity becomes complicated [3, 4]. Margolin et al. first addressed the decomposability into a set of pairwise interactions numerically in a simple three-body problem, but the general condition of the decomposability in many-body problems remains unresolved

[5]. Therefore, in this paper we focus mainly on the decomposability of global information into fundamental pairwise interactions.

3. Local Multivariate Dependence

3.1 Coordinate Systems in the Space of Probability Distributions

How many degrees of freedom are required to specify a probability distribution? We start with a brief overview of the coordinate systems of the space of probability distribution [6, 7]. For the sake of simplicity, suppose that the probability distributions are functions of N variables whose values are discrete and taken to be 0 or 1, that is, binary variables. Then, the number of the degrees of freedom to specify $P(x_1, x_2, \ldots, x_N)$ is $2^N - 1$, where the subtraction of unity is due to the normalization condition $\Sigma_x P(x) = 1$. For $N = 2$ we can choose, for example, $P(0, 0)$, $P(0, 1)$, and $P(1, 0)$ as *coordinates* to specify $P(x_1, x_2)$. The choice of the coordinate system is not unique. In the following, we describe three coordinate systems, whose distinct features are illustrated for $N = 3$ for simplicity.

3.1.1 Marginal Coordinate System

This coordinate system has coordinates given by the marginals $(\eta_1, \eta_2, \eta_3, \eta_{12}, \eta_{23}, \eta_{13}, \eta_{123})$:

$$\eta_\lambda[P] = \sum_x x_\lambda\, P(x), \qquad \lambda \in \{1, 2, 3, 12, 23, 13, 123\}, \tag{10}$$

where we introduce the notation λ to simplify the expression such that $x_{ij} = x_i x_j$ $(i, j \in \{1, 2, 3\})$ and $x_{123} = x_1 x_2 x_3$. These η_λ also correspond to the marginals with the corresponding variables equal to 1 because $x_i = 0$ or 1; for example,

$$\eta_1[P] = \sum_{x_2, x_3} P(x_1, x_2, x_3)\,|_{x_1=1} = P_1(x_1 = 1). \tag{11}$$

If two probability distributions have the same η_1, their first marginals $P_1(x_1)$ and $Q_1(x_1)$ are also the same because of the normalization relation $\sum_{x_1} P_1(x_1) = 1$. Similarly, if two probability distributions have the same η_1, η_2, and η_{12}, the second marginals $P_{12}(x_1, x_2)$ and $Q_{12}(x_1, x_2)$ are also the same. Thus, by using this coordinate system, we can rewrite the definition of the marginal equivalence space in equation (7) as follows:

$$M_\Lambda := \{Q(x)\,|\,\eta_\lambda[Q] = \eta_\lambda[P],\ \lambda \in \Lambda\}. \tag{12}$$

Here Λ is a subset of the combination of $\{1, 2, 3\}$, that is, $\Lambda \subset \{1, 2, 3, 12, 23, 13, 123\}$, and we rewrite the MED as

$$P^\Lambda(x) := \arg\max_{Q \in M_\Lambda} H[Q].$$
(13)

The M_Λ easily enables us to quantify or assign the correlation at each order; for example, for $\Lambda^{(2)} := \{1, 2, 3, 12, 23, 13\}$, $P^{(2)}(x)$ corresponds to the second order MED, implying that the distribution Q in the M_Λ space carries exactly the same information as P for all one body and all pairwise correlations (interactions).

3.1.2 Correlation Coordinate System

This coordinate system is determined by coefficients of the Maclaurin expansion of $\log P(x)$, $\boldsymbol{\theta}[= \{\theta_\lambda\} = (\theta_1, \theta_2, \theta_3, \theta_{12}, \theta_{23}, \theta_{13}, \theta_{123})]$:

$$\log P(x) := \sum_i \theta_i x_i + \sum_{i<j} \theta_{ij} x_i x_j + \theta_{123} x_1 x_2 x_3 - \psi(\boldsymbol{\theta})$$
(14)

where $\psi(\boldsymbol{\theta})$ is the normalization factor.

The condition of $\theta_\lambda = 0$ tells us about the properties of statistical independence of the probability distribution and the maximization of entropy. We discuss the meaning of $\theta_\lambda = 0$ in more detail in Section 3.1.3. Maximizing the entropies yields the following equality [6]:

$$\frac{\partial H[P]}{\partial \eta_\lambda} = -\theta_\lambda.$$
(15)

This equality implies that θ_λ associated with η_λ becomes zero when the probability distribution is determined so as to maximize the entropy with respect to η_λ.

3.1.3 Mixture Coordinate System

As seen in equation (15), η_λ is the marginal, while θ_λ is related to the maximization of the entropies. In the marginal coordinate system $\{\eta_\lambda\}$, it is not easy to naturally incorporate the maximization of the entropies. In turn, the correlation coordinate system $\{\theta_\lambda\}$ may not naturally handle the constraint on the marginals. To resolve this, the following coordinate system is proposed [6]:

$$\left(\eta_{\lambda_1}, \dots, \eta_{\lambda_k}, \theta_{\lambda_{k+1}}, \dots, \theta_{\lambda_7}\right), \qquad 2 \le k \le 7.$$
(16)

For example, if $k = 4$, that is,

$$\Lambda = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\} \subset \{1, 2, 3, 12, 23, 13, 123\},$$

the MED $P^\Lambda(x)$ is represented by the coordinates

$$\left(\eta_{\lambda_1}, \eta_{\lambda_2}, \eta_{\lambda_3}, \eta_{\lambda_4}, \theta_{\lambda_5} = 0, \theta_{\lambda_6} = 0, \theta_{\lambda_7} = 0\right),$$
(17)

where $\lambda_5, \lambda_6, \lambda_7 \in \{1, 2, 3, 12, 23, 13, 123\}\backslash\Lambda$. Since $P^\Lambda(\boldsymbol{x}) \in M_\Lambda$, P^Λ has the marginals $(\eta_{\lambda_1}, \eta_{\lambda_2}, \eta_{\lambda_3}, \eta_{\lambda_4})$, and the other three coordinates are determined by the condition of maximum entropy. By choosing the other three coordinates as $\theta_\lambda = 0$, $\lambda = \lambda_5, \lambda_6, \lambda_7$, equation (17) can be obtained.

This description is not limited to the kth order connected information that takes into account all possible pairs of k-body interactions. For example, if $\Lambda^{12} := \{1, 2, 3, 12\}$, the coordinate of the MED $P^{\Lambda^{12}}$ is given by $(\eta_1, \eta_2, \eta_3, \eta_{12}, \theta_{23} = 0, \theta_{13} = 0, \theta_{123} = 0)$. This distribution has the information of one body and the correlation between x_1 and x_2. Thus,

$$I_C^{12} := \left(-H\left[P^{\Lambda^{12}}\right]\right) - \left(-H\left[P^{\Lambda^{(1)}}\right]\right) \tag{18}$$

(where $\Lambda^{(1)} := \{1, 2, 3\}$) measures the information containing only the correlation between x_1 and x_2. Equation (18) is represented graphically in Figure 1(a). The first order marginals η_i correspond to the vertices, and the second order marginal η_{ij} corresponds to the edge between i and j. The condition $\theta_{ij} = 0$ corresponds to no edge between i and j. By using this graph, the second order connected information is drawn in Figure 1(b).
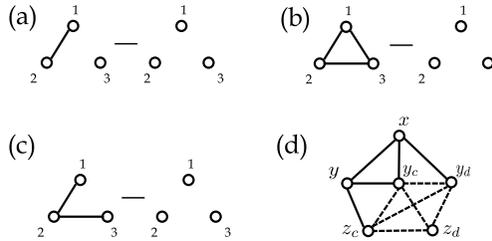

Figure 1. (a) One pairwise interaction information. (b) The second connected information. (c) Two pairwise interactions information. (d) Both solid and dotted lines represent pairwise interactions and the dotted lines are those on which rules 1 and 2 do not depend.

3.2 The Coordinate θ and the Markovian Properties

As discussed earlier, the values of the $\boldsymbol{\theta}$ coordinates relate to the dependence among the variables. In this section, we discuss this relationship in more detail. For the sake of brevity, we consider systems having only pairwise correlations, that is, the third or higher order correlations are absent. In the three-variables case, for example, if x_1 is not connected with x_3, we have the relation

$$P(x_1 \mid x_2, x_3) = P(x_1 \mid x_2), \tag{19}$$

where $P(x \mid y)$ is the conditional probability defined by $P(x \mid y) = P(x, y) / P(y)$. Equation (19) is the Markov process relation, meaning that the information of x_1 depends only on x_2 and the knowledge of x_3 is irrelevant for knowing x_1 (see Figure 1(c)). The derivation of equation (19) is as follows. From the definition of θ in equation (14), the conditions $\theta_{13} = \theta_{123} = 0$ imply that the probability distribution $P(x_1, x_2, x_3)$ can be expressed as a product of some functions $f(x_1, x_2)$ and $g(x_2, x_3)$,

$$P(x_1, x_2, x_3) = f(x_1, x_2)\, g(x_2, x_3). \tag{20}$$

From this expression, we obtain the relation

$$\frac{P_{12}(x_1, x_2)\, P_{23}(x_2, x_3)}{P_2(x_2)} = P(x_1, x_2, x_3), \tag{21}$$

which is equivalent to equation (19).

This Markovian property also applies to cases with more than three variables. For example, see Figure 1(d). There are six variables such that the vertex x connects with only y, y_c, and y_d, and the vertex y connects with only x, y_c, and z_c. The Markovian rules in Figure 1(d) can be stated as

1. $P(x \mid y, y_c, y_d, z_c, z_d) = P(x \mid y, y_c, y_d)$

2. $P(y \mid y_c, y_d, z_c, z_d)$ cannot be reduced to probabilities that do not have variable y_d, even though y does not directly connect with y_d. The variable z_d does not carry any information about y, yielding $P(y \mid y_c, y_d, z_c, z_d) = P(y \mid y_c, y_d, z_c)$.

Note that these rules do not depend on how y_c, y_d, z_c, and z_d are connected with each other, and hence we depicted these possible connections as dotted lines in the figure. Rule 1 means that z_c and z_d give redundant information about x if we know y, y_c, and y_d that connect to x. Rule 2 means that the dependence on the variable y_d in the distribution cannot be dropped because y_d contains the information of x connecting directly to y. In contrast, z_d does not contain the information of x so that variable z_d can be dropped. We can prove these relations in the same way as equation (19).

▎ 3.3 Three-Body Case

As seen so far, the second order connected information, in principle, involves all possible pairwise interactions. We need to establish the underlying relationship between the global connected information and the composite pairwise interactions, or more generally, the local connected information. First, consider the simplest case in which the number of variables is three. Here, the second order connected informa-

tion has all three pairwise interactions. It might be expected that the second order connected information can be expressed as the sum of the individual pairwise interactions, that is,

$$I_C^{(2)} \overset{?}{=} I_C^{12} + I_C^{23} + I_C^{13}. \tag{22}$$

Yet this expectation is incorrect in general. Nevertheless, it can be found that the following decomposition holds:

$$I_C^{12,23} = I_C^{12} + I_C^{23}, \tag{23}$$

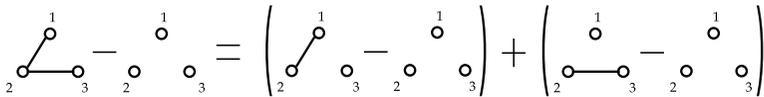whose graphical representation is shown in Figure 2. In what follows we give the proof of equation (23).

Figure 2. Two edges that share one vertex.

From Figure 2, x_1 does not connect with x_3 (i.e., $\theta_{123} = \theta_{13} = 0$). Thus, we obtain the following relation for the MED $P^{\Lambda^{12,23}}$ by using the Markovian relations:

$$P^{\Lambda^{12,23}}(x_1, x_2, x_3) = P^{\Lambda^{12,23}}(x_1 \mid x_2, x_3) \, P_{23}^{\Lambda^{12,23}}(x_2, x_3), \tag{24}$$

$$= P^{\Lambda^{12,23}}(x_1 \mid x_2) \, P_{23}^{\Lambda^{12,23}}(x_2, x_3), \tag{25}$$

$$= \frac{P_{12}^{\Lambda^{12,23}}(x_1, x_2) \, P_{23}^{\Lambda^{12,23}}(x_2, x_3)}{P_2^{\Lambda^{12,23}}(x_2)}. \tag{26}$$

This results in the following relation among the entropies:

$$H\left[P^{\Lambda^{12,23}}\right] = H\left[P_{12}^{\Lambda^{12,23}}\right] + H\left[P_{23}^{\Lambda^{12,23}}\right] - H\left[P_2^{\Lambda^{12,23}}\right], \tag{27}$$

$$= H\left[P_{12}^{\Lambda^{12}}\right] + H\left[P_{23}^{\Lambda^{23}}\right] - H\left[P_2^{\Lambda^{(1)}}\right], \tag{28}$$

$$= H\left[P^{\Lambda^{12}}\right] + H\left[P^{\Lambda^{23}}\right] - H\left[P^{\Lambda^{(1)}}\right], \tag{29}$$

where, for example, $P^{\Lambda^{12}} \overset{\text{def}}{=} P_{12}^{\Lambda^{12}} P_3^{\Lambda^{12}}$ and $P^{\Lambda^{(1)}} \overset{\text{def}}{=} P_1^{\Lambda^{(1)}} P_2^{\Lambda^{(1)}} P_3^{\Lambda^{(1)}}$. The first equality simply arises from equation (26). The second equality follows from the nature of the coordinates of the probability distribution; for example, $P_{12}^{\Lambda^{12,23}}(x_1, x_2)$ has two binary variables, imply-

ing that there are three degrees of freedom that can be specified by the coordinates $(\eta_1, \eta_2, \eta_{12})$. This corresponds to that of the probability distribution $P_{12}^{\Lambda^{12}}(x_1, x_2)$.

This example tells us that the following two necessary conditions must be satisfied for a graph to be decomposable. The first condition is that the vertices are not fully connected with each other so that the conditional probabilities can be reduced to the one with fewer variables (e.g., from equation (24) to equation (25)). The second condition is that the distributions of the fewer variables (e.g., $P_{12}^{\Lambda^{12}}(x_1, x_2)$ in equation (28)) are the same as those with more variables (e.g., $P_{12}^{\Lambda^{12,23}}(x_1, x_2, x_3)$ in equation (27)). In the full graph corresponding to equation (22), the first condition is not met and therefore its global second order connected information cannot be decomposed.

We summarize the given results as follows.

The information of all pairwise interactions cannot be decomposed into those from a subset of pairwise interactions if all variables are fully connected. However, if unconnected pairs exist, we may be able to decompose the global pairwise interactions into a certain set of local interactions.

3.4 Decomposability of Global Connected Information: Graphs That Share One Edge or Vertex

As shown earlier, the global connected information corresponding to the fully connected graphs cannot be decomposed into a set of subgraphs. This does not mean that graphs that are not fully connected can always be decomposed. What kinds of graphs can be decomposed into smaller pieces of subgraphs that enable us to rationalize global features in terms of the composite elements? Here is the answer: if the global connected information corresponding to a graph consists of subgraphs that share only one edge or one vertex with each other, this global connected information can be decomposed into the sum of the local connected information corresponding to these subgraphs, subtracted by the local connected information associated with the sharing edge or vertex of the subgraphs.

This decomposability rule can be generally proven but doing so is beyond the scope of this paper. Here, we simply illustrate this rule by considering a concrete example as shown in Figure 3. The graph in Figure 3 consists of two squares that share only one edge. In this case, the probability distribution can be decomposed as follows:

$$P(x_1, x_2, x_3, x_4, x_5, x_6) = \\ P(x_1, x_2 \mid x_3, x_4, x_5, x_6)\, P(x_3, x_4, x_5, x_6), \tag{30}$$

$$= P(x_1, x_2 \mid x_3, x_4)\, P(x_3, x_4, x_5, x_6), \tag{31}$$

$$= \frac{P(x_1, x_2, x_3, x_4)\, P(x_3, x_4, x_5, x_6)}{P(x_3, x_4)}, \tag{32}$$

where the first equality comes from the chain rule of probabilities and the Markovian relations in Section 3.2 were used in the second equality. Moreover, since the marginal distribution (MD) $P(x_1, x_2, x_3, x_4)$ is equal to $\sum_{x_5, x_6} P(x_1, x_2, x_3, x_4, x_5, x_6)$, its coordinates in the mixture coordinate system are given by

$$\begin{aligned}
&\left(\eta_1, \eta_2, \eta_3, \eta_4, \eta_{12}, \eta_{24}, \eta_{34}, \eta_{13}, \right.\\
&\quad \left. \theta_{14} = 0,\, \theta_{23} = 0,\, \theta_{ijk} = 0,\, \theta_{1234} = 0 \right)
\end{aligned}$$

with $i, j, k \in \{1, 2, 3, 4\}$. These coordinates are the same as those of the probability distribution corresponding to the square subgraph with vertices 1, 2, 3, and 4 in Figure 3. Similarly, the coordinates of the MD $P(x_3, x_4, x_5, x_6)$ match those of the square subgraph with vertices 3, 4, 5, and 6. Finally, the MD $P(x_3, x_4)$ appearing in the denominator of equation (32) has the same marginals $\eta_3, \eta_4, \eta_{34}$ as the subgraph corresponding to the shared edge (the edge connecting vertices 3 and 4) of the two square subgraphs. Therefore, the decomposition of the global connected information associated with the graph with two squares sharing one edge can be represented graphically as shown in Figure 3.
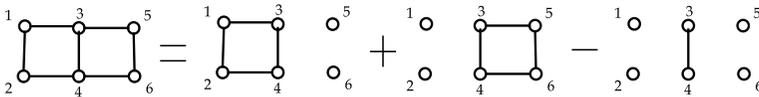

Figure 3. Two squares that share one edge.

3.5 Indecomposable Cases

Here, we look into what circumstances prevent us from decomposing global features into the individual composite elements. Namely, what kinds of graphs that are not fully connected cannot be decomposed into smaller pieces of subgraphs?

Similar to Section 3.1, by using some simple examples as shown in Figure 4, we consider the cases in which two subgraphs share two or more edges.

The first example is a square (see Figure 4(a)) consisting of two sets of two-edges that share two vertices (2 and 3) without any edge connecting them. The MED $P(x_1, x_2, x_3, x_4)$ can be written as follows:

$$P(x_1, x_2, x_3, x_4) = P(x_1 \mid x_2, x_3, x_4)\, P(x_2, x_3, x_4), \tag{33}$$

$$= P(x_1 \mid x_2, x_3)\, P(x_2, x_3, x_4), \tag{34}$$

$$= \frac{P(x_1, x_2, x_3) \, P(x_2, x_3, x_4)}{P(x_2, x_3)}. \tag{35}$$

The MD $P(x_1, x_2, x_3) = \sum_{x_4} P(x_1, x_2, x_3, x_4)$ has the marginal coordinates $\eta_1, \eta_2, \eta_3, \eta_{12}, \eta_{13}$, and $\theta_{123} = 0$. The number of degrees of freedom to specify $P(x_1, x_2, x_3)$ is $7 = 2^3 - 1$ and we must identify the last variable, that is, θ_{23}.

Note that θ_{23} for the MED $P(x_1, x_2, x_3, x_4)$ on the left-hand side of equation (35) vanishes since there is no connection between vertices 2 and 3 in the square graph in Figure 4(a). This can also be seen from the fact (and the Maclaurin expansion, see equation (14)) that $P(x_1, x_2, x_3, x_4)$ can be expressed by a product form of $f_{12} f_{24} f_{34} f_{13}$, where f_{ij} are some functions that depend on x_i and x_j. Therefore, we have

$$\theta_{23} = \frac{\partial^2}{\partial x_2 \, \partial x_3} \log f_{12} f_{24} f_{34} f_{13} |_{x=0} = 0. \tag{36}$$

On the other hand, the coordinate θ_{23} does not vanish, in general, for the MD $P(x_1, x_2, x_3)$ in equation (35) since we have

$$\theta_{23} = \frac{\partial^2}{\partial x_2 \, \partial x_3} \log f_{12} f_{13} \sum_{x_4} (f_{24} f_{34}) |_{x=0} \neq 0. \tag{37}$$

Since $\theta_{23} \neq 0$ for $P(x_1, x_2, x_3)$, the graph corresponding to the local connected information of $P(x_1, x_2, x_3)$ necessarily contains an edge connecting vertices 2 and 3, which is absent in the original square graph. Therefore, the square graph fails to be decomposed into a sum of its subgraphs. We note that the appearance of interaction between x_2 and x_3 (i.e., $\theta_{23} \neq 0$) in the graph corresponding to the MD $P(x_1, x_2, x_3)$ can be regarded as being induced by the simultaneous interactions of $\{x_2, x_1\}$ and $\{x_1, x_3\}$ (and the interactions of $\{x_2, x_4\}$ and $\{x_4, x_3\}$; see Figure 4(a)). The detailed discussion of the properties of these induced interactions will be given elsewhere. The second example in Figure 4(b) consists of two squares that share two edges. Similar to the first example, the MDs $P(x_1, x_3, x_4, x_5)$ and $P(x_2, x_3, x_4, x_5)$ contain induced interaction between x_3 and x_5 that is absent from the original MED, making it impossible to decompose the global connected information.

In the last example shown in Figure 4(c), the graph consists of two full graphs of four vertices that share one triangle. The MED of the global connected information can be expressed using the Markovian relations as

$$P(x_1, x_2, x_3, x_4, x_5) =$$
$$P(x_3 \mid x_1, x_2, x_5) \, P(x_1, x_2, x_4, x_5), \tag{38}$$

$$= \frac{P(x_1, x_2, x_3, x_5) \, P(x_1, x_2, x_4, x_5)}{P(x_1, x_2, x_5)}. \tag{39}$$

Since the subgraphs corresponding to the MDs $P(x_1, x_2, x_3, x_5)$ and $P(x_1, x_2, x_4, x_5)$ are fully connected graphs, there is no induced "pairwise" interaction as in the first and second examples. However, as shown in the following, $P(x_1, x_2, x_3, x_5)$ contains induced "triple" interaction that is absent from the original global connected information. Therefore, the graph in Figure 4(c) cannot be decomposed either. The possible appearance of the induced triple interaction can be understood as follows. The MED $P(x_1, x_2, x_3, x_4, x_5)$ can be expressed using the Maclaurin expansion as a product form $f_{12} f_{13} f_{14} f_{15} f_{23} f_{24} f_{25} f_{35} f_{45}$, where f_{ij} are some functions depending on x_i and x_j. We can also express the marginal probability $P(x_1, x_2, x_3, x_5)$ as $f_{12} f_{13} f_{15} f_{23} f_{25} f_{35} \sum_{x_4} (f_{14} f_{24} f_{45})$, in which the factor involving the summation depends on x_1, x_2, and x_5. The coordinate θ_{125} originating from the Maclaurin expansion of this marginal probability is in general given by

$$\theta_{125} =$$

$$\frac{\partial^3}{\partial x_1 \, \partial x_2 \, \partial x_5} \log \sum_{x_4} P(x_1, x_2, x_3, x_4, x_5) \mid_{x=0} \neq 0. \tag{40}$$

This implies the possible appearance of the induced triple interaction in the graph corresponding to the MD $P(x_1, x_2, x_3, x_5)$. Similar arguments also hold for $P(x_1, x_2, x_4, x_5)$ in equation (39).

The common feature of these three indecomposable cases is that MEDs resulting from the division of the original graph contain apparent interactions the original MED does not possess. Such apparent interactions are regarded as a result of hidden variables. For example, in Figure 4(a) $P(x_1, x_2, x_3)$ is an MD that is obtained by the summation of $P(x_1, x_2, x_3, x_4)$ over x_4. In the original probability distribution x_2 and x_3 are not directly connected, but they are indirectly connected via x_4. The summation of $P(x_1, x_2, x_3, x_4)$ over x_4 inevitably results in a "direct" connection between x_2 and x_3 in $P(x_1, x_2, x_3)$, which originally arises from the indirect connection via the variable x_4 hidden in the MD. In addition, such emergent interactions (e.g., $\theta_{23} \neq 0$ in $P(x_1, x_2, x_3)$) differ from the corresponding interactions in the original joint probability distribution (i.e., η_{23} in $P(x_1, x_2, x_3, x_4)$), because the emergent interactions are not subject

to the marginal in the original distribution. As seen in Figure 4(c), apparent direct interactions arising from hidden variable(s) can also create higher body interactions that the original distribution does not possess. Such apparent direct interactions prevent us from decomposing the global connected information. We can summarize these findings as:

> The existence of interactions due to hidden variable(s) relate to the indecomposability of the connected information.
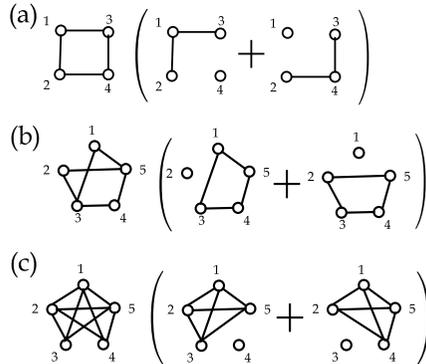

Figure 4. Some indecomposable graphs. (a) One square that shares two vertices. (b) Two squares that share two edges. (c) The five body case with one missing edge, which consists of two full graphs of four variables that share one triangle.

4. Conclusion

By introducing a graphical representation for the multivariate pairwise interacting network and establishing the corresponding Markovian relations, we have formulated the rules for decomposing global pairwise interaction into a sum of local pairwise interactions that cannot be decomposed further (i.e., indecomposable). Since the connected information of the global pairwise interaction is a superposition of those from its indecomposable subsystems, the consequences obtained in this paper are expected to provide us with a decomposition of a complex system (such as a neural network) into mutually independent functional modules or motifs based only on the observation of its probability distributions.

Although systems with binary variables and pairwise interactions are the focus in this paper due to their strong connections to neural systems, the generalization of the given discussion to nonbinary and higher-order interactions is progressing and will be reported separately. On the other hand, it is interesting to extend the current time-independent picture, in which all physical quantities are time inde-

pendent, to the time-dependent regime in which interactions among elements in complex systems can change in time, implying the possibility of decomposing the global interaction into time-dependent functional modules. Such generalization is expected to provide important insights in revealing the evolution of neural systems and in the large scale switching of global neural activities in response to external stimuli.

Acknowledgments

References

[1] E. Schneidman, S. Still, M. J. Berry II, and W. Bialek, "Network Information and Connected Correlations," *Physical Review Letters*, **91**(23), 2003 p. 238701. doi:10.1103/PhysRevLett.91.238701.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[3] E. Schneidman, M. J. Berry II, R. Segev, and W. Bialek, "Weak Pairwise Correlations Imply Strongly Correlated Network States in a Neural Population," *Nature*, **440**, 2006 pp. 1007–1012. doi:10.1038/nature04701.

[4] J. de la Rocha, B. Doiron, E. Shea-Brown, K. Josić, and A. Reyes, "Correlation between Neural Spike Trains Increases with Firing Rate," *Nature*, **448**, 2007 pp. 802–806. doi:10.1038/nature06028.

[5] A. A. Margolin, K. Wang, A. Califano, and I. Nemenman, "Multivariate Dependence and Genetic Networks Inference," *IET Systems Biology*, **4**(6), 2010 pp. 428–440. doi:10.1049/iet-syb.2010.0009.

[6] S. Amari and H. Nagaoka, *Translations of Mathematical Monographs: Method of Information Geometry* (D. Harada, trans.), vol. 191, Providence, RI: American Mathematical Society and Oxford University Press, 2000.

[7] S. Amari, "Information Geometry on Hierarchy of Probability Distributions, *IEEE Transactions on Information Theory*, **47**(5), 2001 pp. 1701–1711. doi:10.1109/18.930911.