Energy Landscape Analysis Toolbox (ELAT) User's Guide (ver. 3.2)

Takahiro Ezaki (https://sites.google.com/site/ezakitakahiro/cv)

Naoki Masuda (http://www.naokimasuda.net)

Latest update: 11 Jan 2022

Table of Contents

Introduction: What is the energy landscape analysis?

Thank you for considering using the energy landscape analysis toolbox (ELAT). The energy landscape (EL) analysis is a computational method that enables intuitive interpretation of multivariate time series. This analysis comprises four steps schematically shown in Fig. 1: (1) Binarization of the data, (2) estimation of the maximum entropy model (a.k.a. Boltzmann machine), (3) construction of a disconnectivity graph and the basin of energy local minima, and (4) computation of indices from the estimated EL. This method was originally designed for analyzing fMRI data, but it is also applicable to other types of data. In our experience, the EL analysis works well when the number of variables is roughly between 6 and 15, depending on the length of the given time series data. In the case of more variables, the computational cost is substantially higher, and the interpretation of the results tends to be difficult. In such cases, we recommend reducing the number of variables (by using ICA, merging variables, etc.).

For detailed algorithms of the EL analysis, we refer to Ref [1]. Example studies are found in Refs. [1-7]. If you have any questions or find errors, please contact Takahiro Ezaki (ezaki0705@gmail.com).

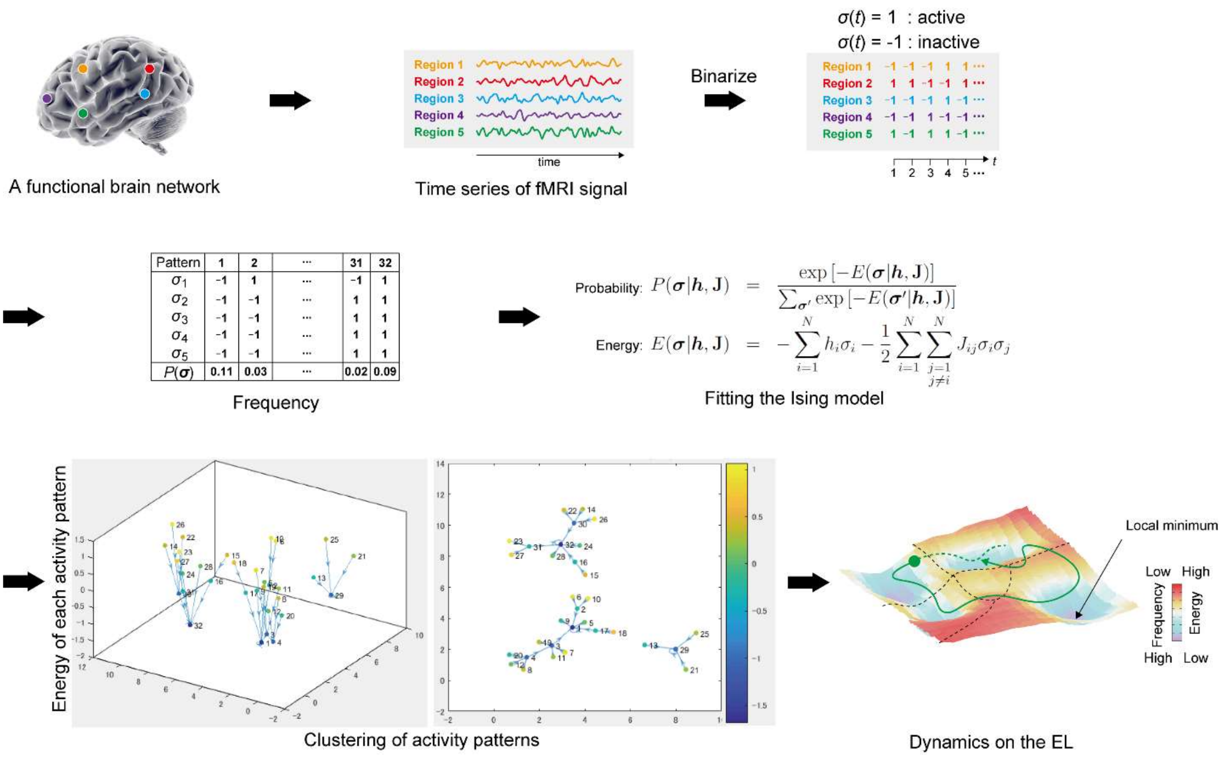

Fig. 1. Overview of the EL analysis.

This toolbox performs the entire computations necessary for the EL analysis (Fig. 2). The Full Analysis option computes everything including the calculation of EL-based indices for each individual. The Energy Landscape Construction option computes the EL and does not perform further analysis.

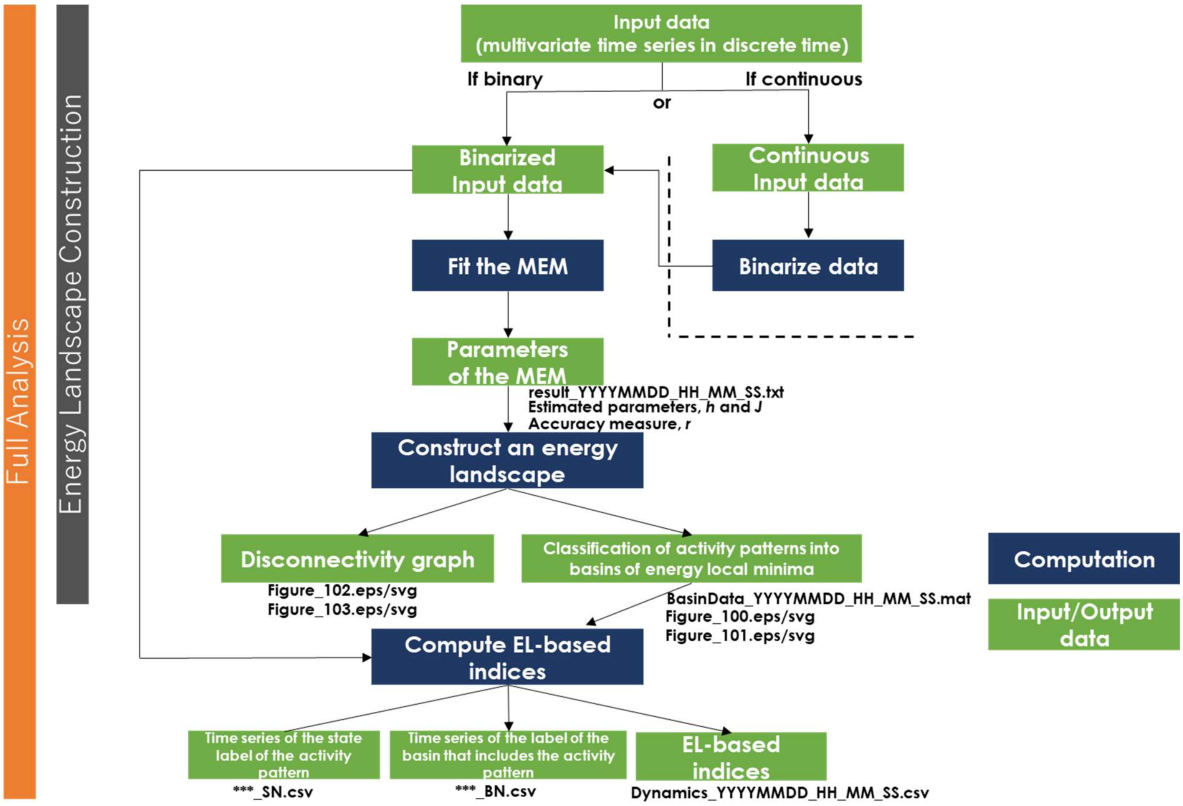

Fig. 2. Overview of ELAT.

Technical details of the EL analysis

We denote the input data by $x_i(t)$, where i ($i = 1, \ldots, N$) labels a variable and t ($t = 1, \ldots, t_{\max}$) represents time.

1. Binarization

The input data for the EL analysis must be binary (i.e., +1 / -1). Thus, if your data $\{x_i(t)\}$ take continuous values, you have to binarize them with an appropriate threshold. In our previous studies [1,7], we binarized our fMRI data by setting a threshold to the average of the signal for each variable, i.e., **$Y_i(t) = 1$ if $x_i(t) > \langle x_i \rangle$, and $Y_i(t) = -1$ otherwise**, where $\langle A \rangle$ denotes the time average of A. Because the baseline of the signal may be different across individuals, we recommend that users separately perform binarization for each individual and each variable. This procedure is implemented in this toolbox and will be explained later.

2. Labeling the activity patterns

The activity pattern of the system at time t is represented by a binary vector $(Y_1(t), Y_2(t), \ldots, Y_N(t))$. This vector is one of the 2^N possible activity patterns. For convenience, we enumerate the activity pattern by a simple conversion from a binary to a decimal number given by $s = 1 + \sum 2^i \ (Y_i + 1) / 2$, which we call a state label. The function $(Y_i + 1) / 2$ maps –1 and 1 to 0 and 1, respectively. This mapping transforms, for example,

(–1, –1, –1, –1) to 1,

(1, 1, 1, 1) to 16, and

(1, –1, 1, –1) to 6.

Thus, the binarized activity patterns are now labeled from 1 to 2^N.

3. Accuracy of fitting

The accuracy of the fitting is measured by r, which is output on the console. The definition of this measure is found in Ref. [1]. An r value close to 1 implies that the MEM fitting is successful. We recommend reporting this value (or a different quantification of the accuracy of fitting) in your work. Note that this value is inevitably small when the data is short [1].

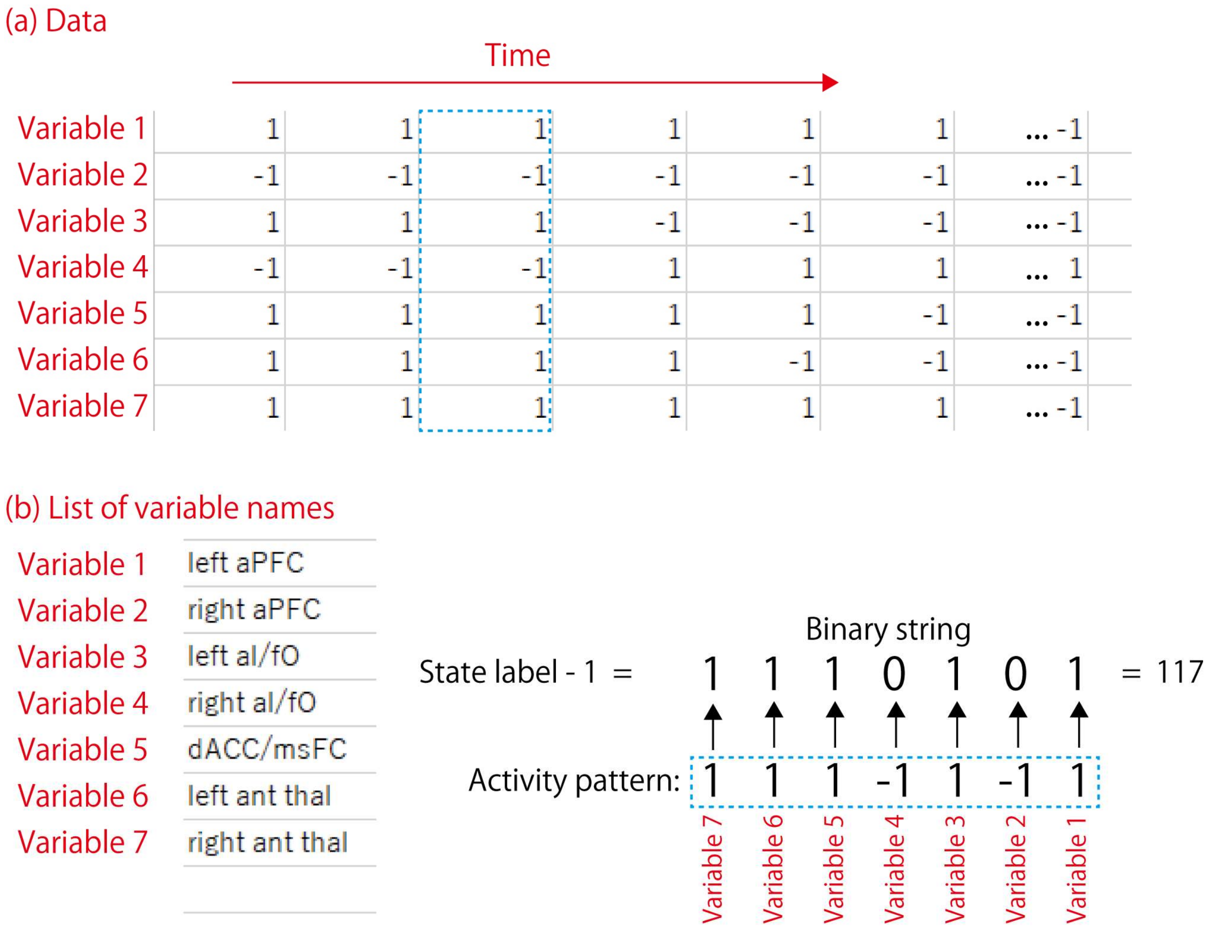

Fig. 3. Input data and the definition of the numerical label of the activity pattern.

How-to 1: A tutorial

Here we show how the toolbox works. (See Fig. 4.)

1. Prepare data

First, prepare the data in a file with extension ".dat". The file must have N (i.e., number of variables) rows and t_{max} columns (see Fig. 3). Each value must be separated by a tab. See "testdata.dat" contained in the toolbox folder for an example. The values in the input data can be either continuous or binarized. If the data is continuous, ELAT binarizes the data by thresholding based on the average of each variable (see Step 5 below). If you do not have data, you can try "testdata_1.dat", …, "testdata_4.dat" in the folder.

2. Launch the toolbox

Unzip the toolbox and place it somewhere convenient for you. Open "StartProgram.m" with MATLAB and run it. Older versions of MATLAB may not support some of the functions used, so please update your MATLAB to the latest version (in many cases, updating is available for free).

3. Select type of analysis

Here, simply select "Full Analysis" (default).

4. Select input data

Select the data you prepared in Step 1. You can select multiple files; each file may correspond to a session of fMRI recording obtained from a participant in the given group, for example. If you select more than one files, ELAT computes a single EL by concatenating the files and then estimating the EL, and also EL indices separately for each file based on the estimated EL. If you want to use the test data, select "testdata_1.dat", …, "testdata_4.dat" in the folder.

5. Select binarization option

Select "Binarized data" or "Continuous data." If you select "Continuous data" and wish to use the average of $\boldsymbol{x}_i$ as the threshold for binarization, please leave the "threshold" 0 (default). If you change this value to c, the threshold will be set to $\boldsymbol{\langle x_i \rangle + c}$. Note that the same $\boldsymbol{c}$ value is applied to all ROIs (i = 1, …, N). If you prefer to binarize the data in a different manner (e.g., with threshold values depending on i), do so on your own and feed the binarized data in Step 4. If you have selected the test data in Step 4, please select "Binarized data."

6. Select Basin Data Generation option

Here, just go with "Construct energy landscape from input files." We will explain the "Read basin data" option later (p. 12).

7. Select variable name (ROI name)

If you have a list of the names of the variables, please select an input file here. The file must be in a ".dat" format and have N rows and 1 column. This step is optional. For an example, see "roiname.dat" in the folder.

8. Select output folder

Select the output folder. We recommend that you create a new folder dedicated to this analysis. If you want to see the list of the EL basins, please also check "Save Basin List."

9. Execute

Execute!

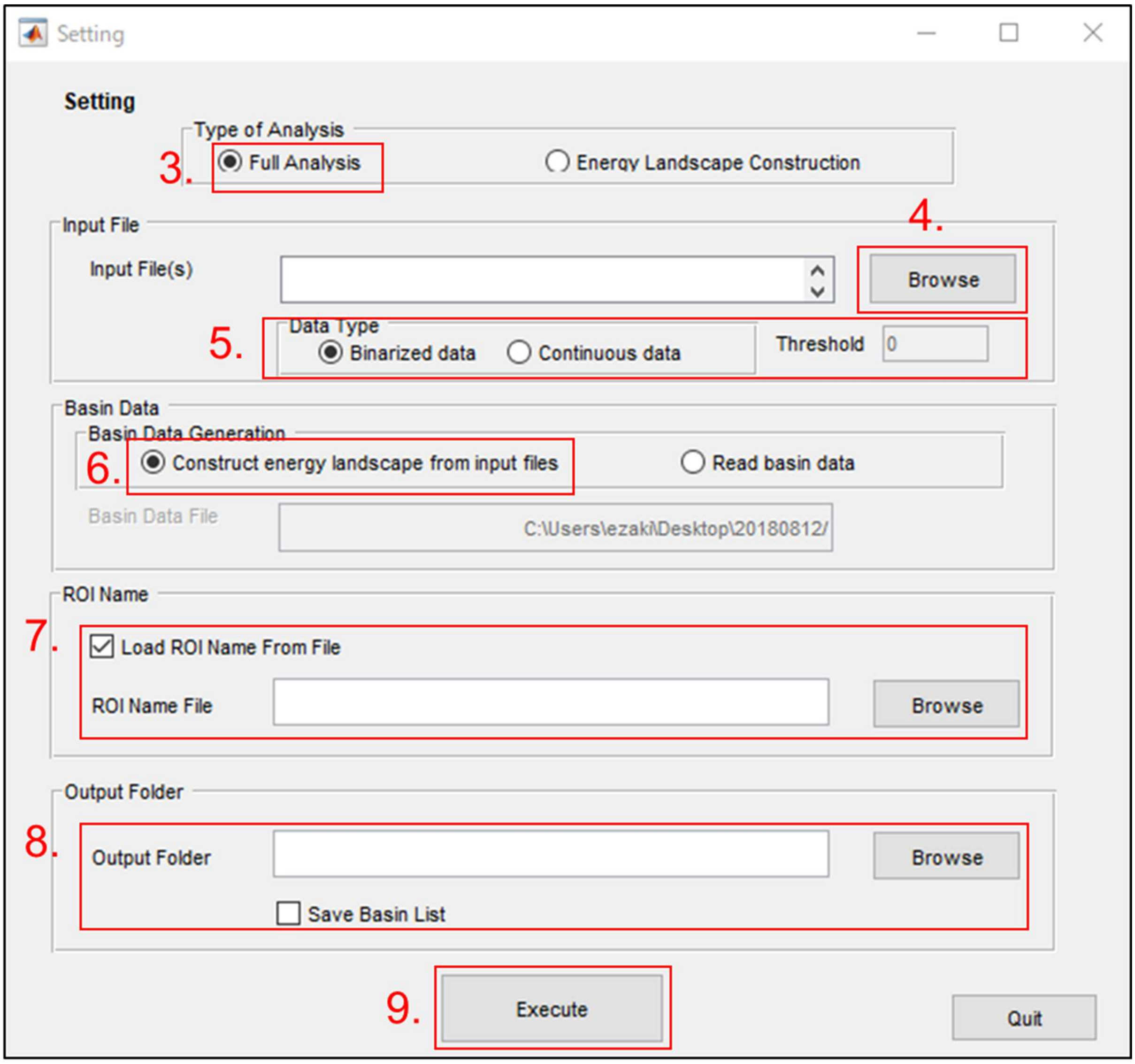

Fig. 4. Configuration window.

How-to 2: Results

1. EL

As a result, first, you will get four figures. The first two figures visualize the EL (Fig. 5). Each node represents an activity pattern specified by its numerical label. Each arrow visualizes the directed edge from an activity pattern to another along the steepest path (in terms of the energy value) toward a local minimum of the energy. The absolute positions of the nodes in the visualization are irrelevant. Each component (i.e., island) of nodes represents the "**basin**" of a local minimum of the energy. The energy value of each activity pattern is shown with a color, which is also visualized by the height in the 3D version shown in the right panel of Fig. 5. The activity patterns with a small energy value appear frequently and may be important.

The two figures are saved in the output folder.

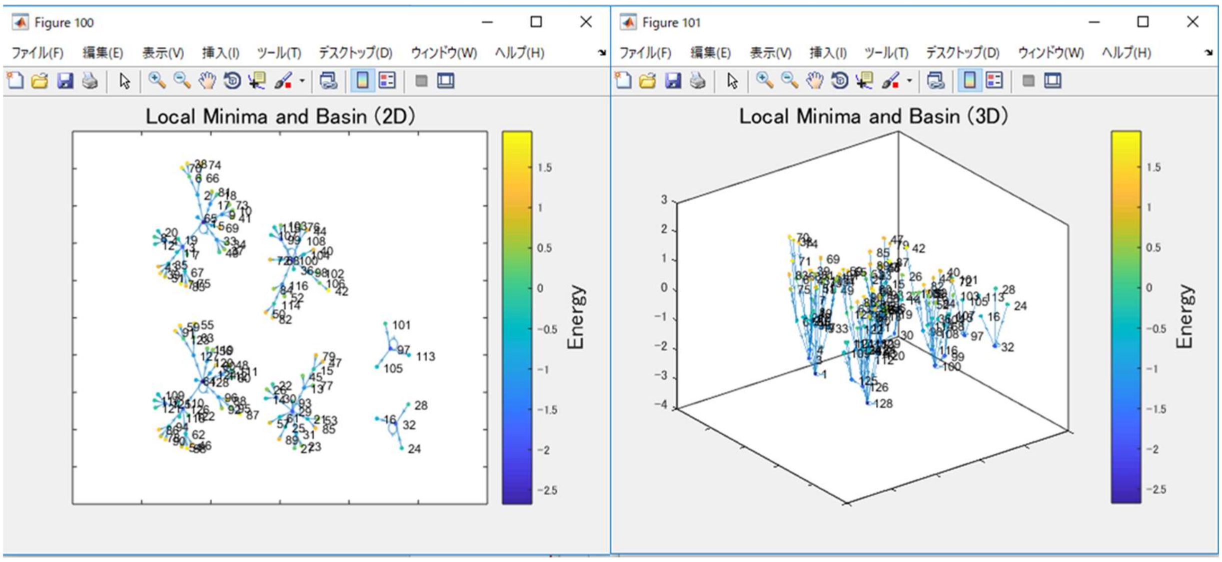

Fig. 5. First two output figures.

The other two figures give a more abstract representation of the EL. The disconnectivity graph is shown in the left panel of Fig. 6. It shows the energy values of all the local minimum activity patterns, which we show in the right panel of Fig. 6, and the height of the energy barriers between them. If you perform this analysis separately for multiple groups of data (e.g., disease group and healthy group), the disconnectivity graphs, one constructed for each group, enable us to compare the different groups. Note that the information about the basin is absent in the disconnectivity graph.

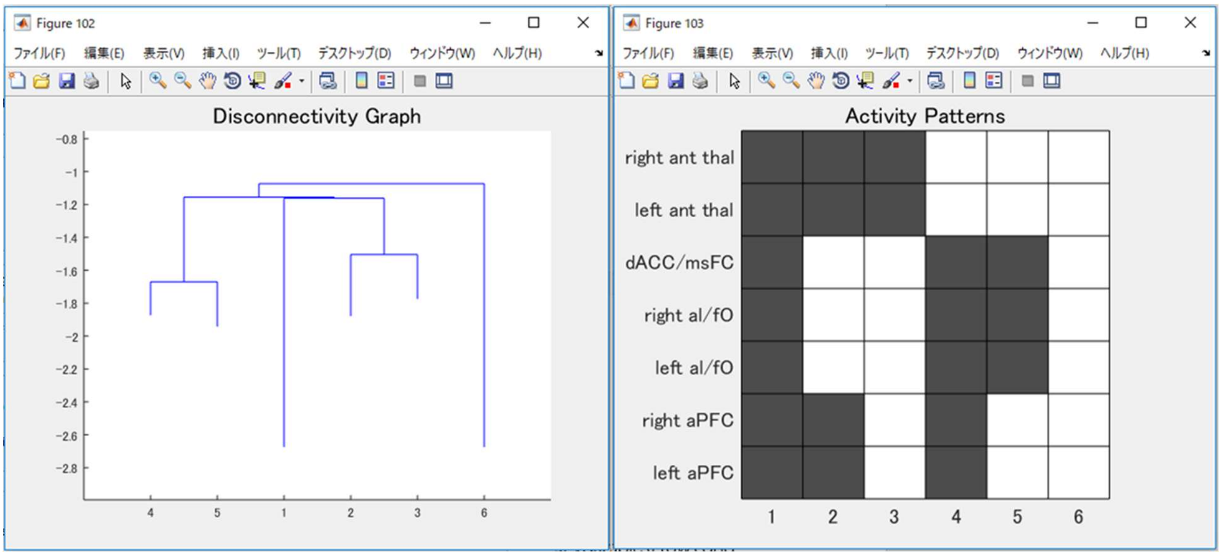

Fig. 6. Disconnectivity graph (left) and activity pattern of each energy local minimum (right). The numbers on the x-axis label the energy local minima, which are consistently used in both panels. In the right panel, a white cell represents an active (i.e., +1) variable, and a gray cell an inactive (i.e., –1) variable. Note that the numerical labels of the energy local minimum activity patterns in this figure are not the same as those used in Figs. 3 and 5.

2. Time series

In the output folder, csv files are created for each input file.

"*_SN.csv" (***: input file name):**

The time series of the activity pattern valued between 1 and 2^N , denoted by ***s(t)***, is saved.

"*_BN.csv":**

The time series of the basin to which ***s(t)*** belongs, denoted by ***b(t)***, is saved. The value represents the label of the basin used in Fig. 6.

3. EL-based indices

In the output folder, a csv file named "Dynamics_yyyymmdd_HH_MM_SS.csv" is created. The indices calculated for each input file.

InputFile:

The name of the data file.

Frequency of *Bj*:

The fraction of time for which ***b(t)*** = ***j***.

Direct transition from *Bj* to *Bk*:

The number of direct transitions from basin ***j*** to basin ***k***, which is divided by t_{max}. Indirect transitions, e.g.,

$j \rightarrow m \rightarrow k$ ($m \neq j, k$)

are excluded.

Transition from *Bj* to *Bk*:

The number of transitions from basin ***j*** to basin ***k***, which is divided by t_{max}. Indirect transitions, e.g.,

$j \rightarrow m \rightarrow k$ ($m \neq j, k$)

are included.

Other options

1. Compute indices on other ELs

In the tutorial, we constructed an EL from the input data and computed indices on it. By choosing the “Read basin data” option, you can compute the indices based on a different EL constructed from other input data. For example, if you have generated an EL using a large data set and found that a specific index is effective at predicting a behavioral measure of an individual, it may be useful to compute the same index for data obtained from new single individuals (e.g., for diagnosis). When you run the Full Analysis, the basin data is saved as “BasinData_YYYYMMDD_HHMMSS.mat” in the output folder.

2. Basin data

“BasinData_YYYYMMDD_HHMMSS.mat” contains a variable called “BasinGraph.” The first column represents the numerical label of the activity pattern. The second column represents the neighboring activity pattern that has the smallest energy value. The connection from the activity pattern on the first column to that on the second column is shown as a directed edge in Fig. 5. The third column is the local minimum activity pattern that the activity pattern belongs to.

References

[1] T. Ezaki, T. Watanabe, M. Ohzeki, and N. Masuda, "Energy landscape analysis of neuroimaging data," Phil. Trans. R. Soc. A 375, 20160287 (2017).

[2] T. Watanabe, N. Masuda, F. Megumi, R. Kanai, G. Rees “Energy landscape and dynamics of brain activity during human bistable perception,” Nat. Commun. 5:4765 (2014).

[3] T. Watanabe, et al. “Energy landscape analysis of resting-state brain networks,” Front. Neuroinform. 8:12 (2014).

[4] A. Ashourvan, S. Gu, M.G. Mattar, J.M. Vettel, D.S. Bassett “The energy landscape underpinning module dynamics in the human brain connectome,” NeuroImage 157:364–380 (2017).

[5] J. Kang, C. Pae, H.J. Park “Energy landscape analysis of the subcortical brain network unravels system properties beneath resting state dynamics,” NeuroImage 149:153–164 (2017).

[6] Watanabe T, Rees G “Brain network dynamics in high-functioning individuals with autism,” Nat. Commun. 8:16048 (2017).

[7] T. Ezaki, M. Sakaki, T. Watanabe, N. Masuda "Age-related changes in the ease of dynamical transitions in human brain activity," Hum. Brain Map. 39, 2673–2688 (2018).