

Mandatory input files to the EA algorithm:

1. File Manager

Usually named as “fileman.txt” contains the basic instructions for the EA such as:

- Number of states per variable, indicated by “P” for which we always use “2” in our Boolean EA code
- Number of variables (or nodes) in the network, indicated by an “N”
- Files to be used by the EA. When no file is used in a given category, the option is leave in blank as {}. Below is a description of the different files to be possibly used:

WT = {"w1.txt","w2.txt"};

This indicates the wildtype time courses to be used; each wildtype time course input is written between quotes. Above indicates two wildtype time courses to be used by the EA, namely w1.txt and w2.txt.

This type of file is MANDATORY.

KO = {(1, "K1_1.txt"),(1,"K1_2.txt"),(2, "K2.txt")};

This indicates the knockout time courses to be used. These is represented by a pair (i, “K.txt”) where i indicates which variable is being knocked out and “K.txt” refers to the file with such knockout time series (i.e. it is expected that the ith column of its courses is the constant zero.

E.g. from above we see three knockout time courses: two knockout outs for variable 1: (1, "K1_1.txt") and (1,"K1_2.txt") as well as knockout time course for variable 2: "(2, "K2.txt").

This type of file is OPTIONAL. Hence if such file is absent then write: KO={};

REV = {"RevMat.txt"};

In here we input the “Previously applied” Reverse engineering Matrix: This n x (n+1) matrix is the probability-adjacency matrix where for entry a_{ij} represents the

probability between 0 and 1 that an edge from x_j to x_i exists for $i=1,\dots,n$. Column $n+1$ can be left just as constant column 1.

This matrix is obtained from another given reverse engineering method applied previously in data of the network of our interest.

This type of file is OPTIONAL. Hence if such file is absent then write: $REV=\{\}$;

$CMPLX = \{\}$;

Leave this in blank.

$BIO = \{"Bio.txt"\}$;

In here it is input the Biological Matrix: The $n \times n$ matrix is the probability-adjacency matrix where for entry a_{ij} represents the probability between 0 and 1 that an edge from x_j to x_i exists.

This matrix is obtained from information from any previous biological information on the structure of the static network

This type of file is OPTIONAL.

$MODEL = \{"Model1.txt", "Model2.txt"\}$;

In here we input the different files that contain the different “seeding” polynomial models from a previously applied reverse engineering method that returns a polynomial model, such as Stigler and Laubenbacher method.

This type of file is OPTIONAL.

$PARAMS = \{"params1.txt"\}$;

In here we indicate the name of the file that contains the parameters of the EA.

This type of file is MANDATORY.

2. Time Courses File(s)

There are two types of parameter files

- Wildtype
- Knockout (the i th knockout time series is characterized by having its i th column equal to constant zero)

3. Parameter File

Fixed set of parameters Set of Parameters:

MinGenePoolSize. Smallest number of models under test at one time - (10 to MaxGenePoolSize - 1)

MaxGenePoolSize. Maximum number of models under test at one time - (MinGenePoolSize to 5000)

MinParents. Minimum number of parents used as the basis for a next generation (2 - MinGenePoolSize)

NumParentsCopied. Number of the best model copied unchanged to the next generation (1 - MinParents) • MaxGenerations. Number of times to create a new generation (at least 2)

MutateProbability. The probability of changing 1 bit in any polynomial. The probability of changing n bits is p^n . (0.0 to 0.99...)

CrossoverProbability. The probability of switching parents as a child is produced by copying terms from one of the parents. (0.0 to 0.99...)

HammingWeight. Relative significance of the total hamming distance between the time series generated by a model and the corresponding input time series (0.0 to 1.0)

ComplexityWeight. Relative significance of the model complexity (total number of variables in all terms). (Should be 1.0 - HammingWeight)

HammingSumDecay. Used when combining Hamming distances from time series generated with mutated initial states for a single model. The hamming distances are

sorted in increasing order, and then a weighted average is computed where the smaller distances receive more weight. This parameter controls how fast the weights decrease (as $e^{-k \cdot i}$), where i is the index in the table of weights and k is HammingSumDecay. (1 to 10)

MutateInitialState. If not zero then evaluate each model by generating time series based on mutated initial states. If zero, generate time series only using the input initial state.

MaxStates. Maximum number of states of a variable

MaxVariables. Number of different variables in the polynomial model

Optional input files to the EA algorithm:

1. Biological Matrix
2. Previously applied Reverse Engineering Method Matrix
3. Seeding Models

To run EA:

Place in a folder both, mandatory files and EA executable.

To start to run on the command line write: name of executable, follow by the file manager name and indicate a name of file to be created by the EA where the information about the EA computations will be saved.

Example at the command line write:

EA_Two fileman00001.txt > 00001.txt

Above indicates to use EA executable called “EA_Two” and to use the file manager called “fileman00001.txt” and finally, that the information about the computations of the run must be saved in the folder in a file called “00001.txt”.

Additionally to the indicated file “00001.txt”, the EA will return as output the files:

“BestModels.txt” that will give the 35 highest scored models obtained by the EA in its last generation.

“DebugModels.txt” if the debug version of the EA is being used, which returns the best 35 models of each single generation computed by the EA.