Readme file for Zodiac software

The current BGM software can be used on Linux, Mac OS and Windows (32bit and 64bit) system. We provide the binary R-package to run the BGM algorithm. Note that ZodiacBGM is built under R-version 3.0.2. So you need R-version 3.0.2 or later.

Once you download the binary package which is appropriate for your system, you need to install the binary package through the function *install.packages()*. See details of this function call `https://stat.ethz.ch/R-manual/R-devel/library/utils/html/install.packages.html`.

Start R and set the example code folder as the Present Working Directory of R. Read and execute the example code given below to perform BGM analysis on an example dataset (included in the package), which includes gene expression (GE), DNA copy number (CN), and DNA methylation (Me) data of two genes (i.e. A1CF and A2M) from 1448 patient tumors over 11 cancer types. The data is a 6 by 1448 numeric matrix including 6 features/nodes and 1448 samples.

Comments in the example code explain how to organize data and run the network analysis function, which is called *NetworkAnalysis*.

The input to *NetworkAnalysis* is a list variable consisting of four elements:

- **y :** the input data matrix. Each row corresponds to a node in the network and each column corresponds to a sample.
- **Rep :** a positive integer indicating the number of MCMC iterations to be executed. Rep must be no smaller than 10000.
- **graph.prior :** an N by N symmetric 0, 1 numeric matrix representing the prior network used to guide model training, where 1 indicates an interaction and 0 indicates no interaction. N is the number of nodes in the analysis. For example, if the first node and the second node interact, the element in the first row and second column and the element in the second row and first column are 1; otherwise they are 0.
- **pho :** a value used to control the strength of prior network.

The output of NetworkAnalysis is a list variable consisting of five elements:

- **post.mean.beta :** an N by N symmetric, numeric matrix gives the mean values of β parameters sampled in the MCMC chain.
- **post.sd.beta :** an N by N symmetric, numeric matrix gives the standard deviations of β parameters sampled in the MCMC chain.
- **prob.edge :** an N by N symmetric, numeric matrix gives the posterior probabilities of interactions. For example, 0.8 at the first row and second column indicates that the interaction involving the first and second nodes occurs with a posterior probability of 0.8.
- **FDR :** an N by N symmetric, numeric matrix gives the FDR controls of interactions. graphMatrixList a list of no more than 3 elements. It shows the top three most frequent graphs sampled by the MCMC chain, with the element names showing the graph frequencies. Each element is an N by N symmetric 0, 1 numeric matrix representing a graph, where 1 indicates an interaction and 0 indicates no interaction.

For details of the method and examples please refer our paper [1].

Example

```
> # The following is an example of running BGM analysis on an example data
>
> # Clear the workspace
> rm(list = ls());
> # Load ZodiacBGM library
> library(ZodiacBGM)
> # Load example data of A1CF and A2M.
> # There are 6 features/nodes and 1448 samples.
> data(exampleBGM_data);
> # Generate a list called Input to be input into the
> #NetworkAnalysis function for doing BGM analysis;
>
> Input = list(y = exampleBGM_data,
+                 # y is the input data, each row is a node in
+                 # the network and each column is a sample
+                 Rep = 20000,
+                   graph.prior = rbind(c(0, 1, 0, 0, 0, 0), c(1, 0, 1, 0, 0, 0),
+                                      c(0, 1, 0, 0, 0, 0), c(0, 0, 0, 0, 1, 0),
+                                      c(0, 0, 0, 1, 0, 1), c(0, 0, 0, 0, 1, 0)),
+                 # graph.prior is a symmetric 0,1 matrix representing
+                 # the prior network used to guide model training,
+                 # where 1 indicates an interaction and
+                 # 0 indicates no interaction.
+                   pho = 1);
>                 # pho is the pho value used to control
>                 # the strength of prior network
>
> # Run the main function NetworkAnalysis
> Result = NetworkAnalysis(Input);
> # result is shown at the output console
> print("Network Analysis is done !!");

[1] "Network Analysis is done !!"
```

References

[1] Yitan Zhu, Yanxun Xu, Donald L Helseth, Kamalakar Gulukota, Shengjie Yang, Lorenzo L Pesce, Riten Mitra, Peter Müller, Subhajit Sengupta, Wentian Guo, et al. Zodiac: A comprehensive depiction of genetic interactions in cancer by integrating tcga data. *Journal of the National Cancer Institute*, 107(8):djv129, 2015.