**GRNInfer testing on Linux OS report**

**OS**

x86\_64 GNU/Linux

java version "1.7.0\_79"

OpenJDK Runtime Environment (IcedTea 2.5.5) (7u79-2.5.5-1~deb8u1) OpenJDK 64-Bit Server VM (build 24.79-b02, mixed mode)

**Build from source and Testing**

1. Fetching source from <https://bitbucket.org/birc_ntu/metagrn> pass
2. Building from source under eclipse

pass

2. Testing

1. About pass
2. GRN inference 2.2.1 DBN
   1. data\_10G30T\_dbn.txt pass
   2. data\_10G30T\_dbn\_tab.txt pass
   3. Test – Input with header indicating gene names DBN -> “Save result”

pass

DBN -> “visulize result”

the gene name labels wrong.

Liuxingliang2015Jul05: fixed

Read the saved file

pass

(c1) Column header / row header is confusing

If I put the first row as the gene name and then select Column header, it gives the right network. But actually the row header is the gene name.

Please correct me if I am wrong.

Liuxingliang2015Jul05: if the first row is gene name, I think it means that column headers are gene names, just like the following example:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Gene1 | Gene2 | Gene3 |
| Time1 |  |  |  |
| Time2 |  |  |  |
| Time3 |  |  |  |

What is the header of first column? Gene1

What is the header of second column? Gene2

What is the header of second row? Time2

If it is confusing, w can put the input format in our manual to tell users what the column header and row header mean.

* 1. Test – Include prior data

Here I took “user\_network” as Prior data file

The progress bar do not progress

Liuxingliang2015Jul05: your input is not prior data, the prior data should be some epigenetic data of histone modifications, for details, please check Haifen and Kasun’s paper: “Integrating Epigenetic Prior in Dynamic Bayesian Network for Gene Regulatory Network Inference”

* 1. Test – Number of iterations

pass

1. RF
   1. Test RF-Random Forests vs. Extra Trees Result (data5G10T\_en.txt) The inferred network is different

Liuxingliang2015Jul05: I think they should be different, because the algorithm is based on randomization. I think I have already tested them using a fixed random number series. The results of my conversion and original code are same.

|  |  |  |
| --- | --- | --- |
| G1 | G3 | 1 |
| G3 | G1 | 1 |
| G3 | G2 | 1 |
| G3 | G4 | 1 |
| G3 | G5 | 1 |
| G4 | G2 | 1 |

|  |  |  |
| --- | --- | --- |
| G4 | G3 | 1 |
| G4 | G5 | 1 |
| G5 | G1 | 1 |
| G5 | G4 | 1 |
| G1 | G2 | 1 |
| G1 | G3 | 1 |
| G2 | G5 | 1 |
| G3 | G1 | 1 |
| G3 | G2 | 1 |
| G3 | G4 | 1 |
| G4 | G3 | 1 |
| G4 | G5 | 1 |
| G5 | G1 | 1 |
| G5 | G4 | 1 |

1. ElasticNet Basic testing

pass Advanced testing

see below General problems

1. Lasso

Basic testing with 50G, the result has something as below:

|  |  |  |
| --- | --- | --- |
| G6 | G6 | 1 |
| G24 | G24 | 1 |
| G29 | G29 | 1 |
|  |  |  |

Liuxingliang2015Jul05: I think our network is by default with such edges (self-circle).

2.2.5 Time-delayed Lasso

Once the running is finshed, reached 100%. An undesirable window poping up to save the Time-delay result before the dialogue box showing up to ask whether we want to “Save the network” or “Visualize the network.”

Liuxingliang2015Jul05: removed the middle step of asking users to save time-delay information of inferred network, it will be saved as another file named as <inferrednetworkname>\_timedelay.tsv along with the inferred network file

|  |  |  |
| --- | --- | --- |
| 2.2.6 Ridge Regression |  |  |
| The inferred network also contains G2 | G2 | 1 similar cases. |

Liuxingliang2015Jul05: same as above reply

1. Meta GRN (PATH issue)

choose data5G10T\_en.txt, with DBN as the algorithm to compare, it shows me below when I clicked Meta-GRN button.

WARNING: Simulation::run(): write: /DBN/0/DBN\_goldstandard\_signed.tsv (No such file or directory)

java.lang.Exception: write: /DBN/0/DBN\_goldstandard\_signed.tsv (No such file or directory)

additionally, java.lang.Exception: URL has no content / file not found: /scratch/lina/java\_projects/metaGRN/RF.tsv

Liuxingliang2015Jul05: fixed, you use it in a wrong way, I simplified the workflow.

4. Help

Empty

Liuxingliang2015Jul05: since we have manual, so I removed this button.

.

5. Exit

pass

General Problems:

1] Full path of the input file. If I input the absolute path of the file wrong, I can't resume working by selecting another file without closing the window first.

Liuxingliang2015Jul08: fixed. Adding input format check, and other checks to very component which requires input file, if user input is not correct, will promote a message and allow them input again.

If I am not wrong, there are two type of input for our software: data matrix and network file.

For data matrix: the file must be separated by <TAB>,

if it has header information, the format must confirm the following one:

<TAB>gene1<TAB>gene2<TAB>gene3

Time1<TAB>val1<TAB>val2<TAB>val3

Time2<TAB>val1<TAB>val2<TAB>val3

Time3<TAB>val1<TAB>val2<TAB>val3

or, interchange the columns and the rows (matrix transpose)

if it doesn’t have header information, the format doesn’t change, only remove header row and header column.

(Why we choose this type of format is because it is very easy to compose using Excel-like software in Windows OS or using shell in Unix-like system)

For network: we use standard format in GeneNetWeaver. The file format must follow the following one:

Node1<TAB>Node1<TAB>0

Node1<TAB>Node2<TAB>1

Node1<TAB>Node3<TAB>1

Node2<TAB>Node1<TAB>0

Node2<TAB>Node2<TAB>1

Node2<TAB>Node3<TAB>1

Node3<TAB>Node1<TAB>0

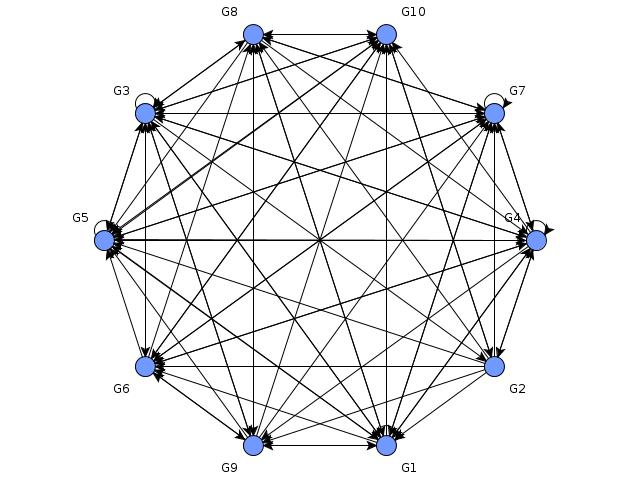
Node3<TAB>Node2<TAB>1

Node3<TAB>Node3<TAB>1

There should be no header, the first two columns are node name (e.g. gene name), and the third column is 0 or 1, which means not connects or connects separately. Please note that, in order to keep format uniformly, we don’t remove any 0 items, duplicates (i.e, node1 -> node2 and node2 -> node1 in no-direction graph) and self-circle (i.e., node1->node1).

2] Long time scale expression data. Here I use 10 Genes with 300T, the DBN, RF gives a reasonable result, the ElasticNet do not degenrate the edge, see below figure.

Liuxingliang2015Jul05: I am not sure it is a normal behavior for EN or not. However, I have already compared the output of my EN code with original one, their output are same.



Liuxingliang2015Jul11:

1. Remove and refine some confusing GUI logic, like network format option in MetaGRN panel and NetworkComparison panel, now we don’t provide matrix format network input, only accept the GNW standard, three-columns, format, like the following one:

Node1<TAB>Node1<TAB>0

Node1<TAB>Node2<TAB>1

Node1<TAB>Node3<TAB>1

Node2<TAB>Node1<TAB>0

Node2<TAB>Node2<TAB>1

Node2<TAB>Node3<TAB>1

Node3<TAB>Node1<TAB>0

Node3<TAB>Node2<TAB>1

Node3<TAB>Node3<TAB>1