tinge

Name

`TINGe` — Tool Inferring Networks of GEnes

Synopsis

tinge-mi {-i *file*} *options*...

Description

TINGe is a parallel tool for constructing gene regulatory networks from large-scale gene expression data. It uses information theoretic criteria and statistical testing to detect dependencies between genes, and it can be run on large parallel machines, such as IBM Blue Gene. TINGe is implemented in C++ and MPI, therefore it should be executed in the MPI environment by using an adequate **mpiexec** command.

Options

In general, TINGe tries to maintain compatibility of options, and file formats, with ARACNe. Please refer to [1] for the detailed description of file formats.

-i *file*

Read expression profiles from `file`.

-o *file*

Write output network to `file`. If option is not specified, the output file is created by changing extension of the input file to `.adj`.

-j *file*

Read relevance network from `file`.

-w *file*

Store relevance network in `file`.

-l *file*

Read list of transcription factors from `file`.

-a *estimator*

Set the estimator that should be used to calculate mutual information to *estimator*. Currently two estimators are provided: *B* for B-spline estimator [2], and *G* for Gaussian kernel estimator [3]. We

highly recommend using the B-spline estimator for its high performance and accuracy. By default the B-spline estimator is used.

-b *bins*

Set the number of bins used by the B-Spline mutual information estimator to *bins*. This option is ignored if Gaussian kernel estimator is used. By default 10 bins are used. See [2] for a more detailed explanation.

-k *order*

Set the order of basis B-Spline functions used by the mutual information estimator to *order*. This option is ignored if Gaussian kernel estimator is used. Default is 4. See [2] for a more detailed explanation.

-p *pval*

Set the significance level of testing if mutual information is statistically greater than 0 to *pval*. This option is mutually exclusive with *-t*. Default is 1.0, that is, no test is performed. See [4] for a more detailed explanation.

-t *threshold*

Set the threshold value for mutual information to *threshold*. This option is mutually exclusive with *-p*.

-e *tolerance*

Set the tolerance of data processing inequality to *tolerance*. Default is 1.0, that is, no processing is applied. See [1] for a more detailed explanation.

-C *method*

Convert final mutual information using *method*. *C* converts mutual information to correlation coefficient using formula by Joe [5]. Currently only this method is supported. By default no conversion is applied.

-r *size*

Set the number of generated bootstrap networks to *size*. This option is mutually exclusive with *-y*. Default is 0, that is, no bootstrapping is applied.

-y *size*

Set the number of columns used for analysis to *size*. If set, only *size* randomly selected columns (with no replacement) are used. This option is mutually exclusive with *-r*. Default is 0, that is, no sampling is applied.

-x *seed*

Set random seed utilized in statistical tests to *seed*. By default seed is generated from the current time and process identifier.

-m

Turn on memory usage reports. By default reports are disabled. On some architectures memory reports might be incorrect.

-v

Show progress, i.e. notify about iteration progress during the main processing phase.

-h

Show summary of options.

Example

```
mpiexec -np 32 tinge-mi -i tost.exp -p 0.001 -e 0.1
```

The above command will execute tinge-mi on 32 processors. A network will be created based on the expression data from `tost.exp` and it will be stored in `tost.adj`. Statistical testing at significance level of 0.001 will be performed, and data processing inequality will be applied with tolerance of 0.1.

References

[1] Margolin, A.A. et al., "Reverse engineering cellular networks", Nature Protocols, vol. 1 no. 2, 2006.

[2] Daub, C.O. et al., "Estimating mutual information using B-spline functions - an improved similarity measure for analysing gene expression data", BMC Bioinformatics, vol. 5 no. 118, 2004.

[3] Moon, Y., Rajagopalan, B., Lall, U., "Estimation of mutual information using kernel density estimators", Physical Review E., vol. 52 no. 3, 1995.

[4] Zola, J., Aluru, M., Sarje, A., Aluru, S., "Parallel information theory based construction of genome wide gene regulatory networks", In IEEE Transactions on Parallel and Distributed Systems, vol. 21, no. 12, pp. 1721-1733, 2010.

[5] Joe, H., "Relative entropy measures of multivariate dependence", Journal of the American Statistical Association, vol. 84 no. 405, 1989.

Bugs

Bugs? What bugs? Well, if you are *sure* that you have found a bug you can contact Jaroslaw Zola.

Author

Jaroslaw Zola <jaroslaw.zola@gmail.com>, Maneesha Aluru, Dan Nettleton, Srinivas Aluru.

Copyright