BinAligner
User's Manual

November 21, 2012

1 Methods

BinAligner is a bioinformatics software for aligning biological networks, which uses a mixture of global and local algorithm for network alignment. BinAligner constructs a pairwise similarity matrix between two networks based on three types of similarity scores derived from vertices (e.g. single node comparison based on sequence similarity), 1-neighbor alignment (e.g. the similarity of two nodes based on the similarity of their first neighbor subnetworks) and graphlet (e.g. the similarity of n-neighborhood subnetworks, $n \geq 2$), which integrate information from both nodes and edges. Then the alignment problem can be formulated as an assignment problem, which can be solved by the combinatorial optimization algorithm in polynomial time.

2 Installation

2.1 Prerequisites

BinAligner requires C++ compiler and CPLEX Optimization Studio:
    - C++ compiler
    - IBM CPLEX Optimization Studio [12.4], which is available to academics through the IBM Academic Initiative.

2.2 Install

The installation of BinAligner consists of two parts:

1. Install IBM CPLEX;

2. Install BinALigner (OS requirement: UNIX/Linux/MaxOSX).

Please read "INSTALL" located in "BinAligner/doc" for the install instructions.

3 Example

After installing the BinAligner, users can try one example by running the command "./run.sh" in directory "example". Before using "run.sh", you should first change its permission to be executable by command "chmod +x run.sh". All result files will be generated in the directory "result".

4 Usage

BinAligner requires 5 input files: 2 network files, 2 node files and 1 node similarity file. All the other parameters are optional. BinAligner will use their default values if they are not specified.

To use, run the following command:

./binaligner [PARAMS]

PARAMS:

```
-m, --net1-file        network 1 data file.
-n, --net2-file        network 2 data file.
-i, --node1-file       node labels for network 1.
-j, --node2-file       node labels for network 2.
-s, --orth-file        node similarities between two networks.
-o, --output-dir       output directory
-r, --result-file      result summary file
-k, --graph-file       alignment graph file
-p, --alignment-file   alignment file
-l, --lambda           parameter lambda.
-g, --graphlet-coef    the weight of graphlet similarity matrix.
-f, --orth-coef        the weight of vertex similarity matrix.
-a, --alpha            edge matching score
-b, --beta             mis-match/non-match penalty score
-c, --gamma            vertex ortholog score
-d, --delta            vertex non-ortholog score
-h, --help             help for the parameters and their usage.
-v, --verbose
```

The shell script "run.sh" in the directory "BinAligner/example" shows how to use BinAligner to run the example data.

5 Visualization

The final alignment is a GraphViz .dot file, which can be visualized by GraphViz.

6 File format

6.1 Input file format

BinAligner takes three types of input files: (i) node file, containing all the node labels. (ii) network file, containing all the links of the network. (ii) ortholog file, containing all the ortholog pairs.

6.1.1 Node file

A sample node file (each line represents a node):

```
K10
K15
K3
K5
K7
K9
Orf2
Orf23
Orf28
Orf30
Orf31
Orf39
Orf41
Orf45
Orf47
Orf52
Orf6
Orf67.5
Orf68
Orf74
Orf9
```

6.1.2 Network file

A sample network file (each line represents an edge, two nodes are separated by a single space character):

```
Orf6 K5
Orf6 Orf52
Orf6 K15
Orf9 Orf39
Orf9 Orf41
Orf9 Orf47
Orf9 K10
Orf9 Orf67.5
Orf9 Orf68
K10 Orf39
K10 Orf41
K10 Orf47
K10 Orf67.5
K10 Orf68
K10 Orf31
Orf2 K10
K7 K5
K7 K3
K7 Orf74
Orf23 Orf67.5
Orf23 Orf28
Orf23 Orf30
Orf23 Orf45
Orf23 K9
Orf28 K5
Orf28 Orf41
Orf28 K10
Orf28 Orf67.5
```

6.1.3 Ortholog file

Here is a sample ortholog file (space separated). Each line represents a node pair, which the first one is in network1 and the second is in network2.

```
Orf9 Orf28
Orf39 Orf50
Orf67.5 Orf25
Orf68 Orf26
Orf29b Orf42
Orf61 Orf19
Orf60 Orf18
Orf69 Orf27
Orf74 Orf36
```

```
Orf52 Orf1
Orf28 Orf1
Orf53 Orf1
Orf52 Orf46
Orf67.5 Orf7
Orf72 Orf7
Orf67.5 Orf9a
Orf30 Orf9a
Orf53 Orf9a
K8 Orf23
K15 Orf65
Orf28 Orf65
Orf53 Orf65
Orf29b Orf45
Orf67.5 Orf49
Orf30 Orf57
```

6.2 Output file format

6.2.1 Result summary file

Here is an example of the result summary file. First three columns are the basic parameters used. The last four columns are the number of aligned nodes, aligned edges, ortholog pairs and loops.

```
lambda  graphlet_coef    orth_coef    #nodes  #edges   #orths  #loops
1       0.01             0.9          40      40       16      0
```

6.2.2 Alignment file

As shown below, each row in the file represents an aligned node pair. Two nodes in an aligned pair are ortholog if "ortholog" is 1.

```
node1   node2      ortholog
Orf27   S/L           0
Orf28   Orf1          1
Orf29b  Orf42         1
Orf30   Orf57         1
Orf63   Orf33         0
Orf65   Orf56         0
```

6.2.3 Graph file

There are three types of colored edges between two vertices: (i) red edges, representing the edge matches existing in both network 1 and network 2; (ii) blue edges, the edges only in

network 1; (iii) green edges, the edges only in network 2. Each vertex represents a pair of aligned vertices. The shaded pink vertices are aligned orthologous pairs. The graph file (with .dot) generated by BinAligner can be visualized by GraphViz.

7   Legal

- BinAligner is released under the GNU General Public License as set down at

<div align="center">

http://www.gnu.org/copyleft/gpl.html

</div>

References

[1] Yang J, Li J, Grünewald S, Wan X. BinAligner: a heuristic method to align biological networks. Manuscript 2012.