NetworkEvolution plugin User's Manual

Release 2.1

Michał Woźniak, Janusz Dutkowski, Jerzy Tiuryn

05 March 2010

Contents

1 About

1.1 About plugin

The *NetworkEvolution* plugin is a tool for interactive comparative analysis of networks across different species. Our primary goal here is to allow the user to walk step-by-step through the process of evolution of a selected network module. This process possibly includes protein duplications, losses and speciation, as well as the emergence or deletion of network links. Parent-child and sibling-sibling relationships between homologous proteins are also taken into account and color-coded to allow visual interpretation. Standard input data is provided by the CAPPI procedure (Dutkowski and Tiuryn, Bioinformatics 2007). The data includes the species tree, protein family trees, reconstructed interaction networks and experimentally identified interactions.

It is an open source application (GPL 3 license).

The plugin jar file, instructions and sources are available at:

`http://bioputer.mimuw.edu.pl/modevo`

Sources are available at code.google website:

`http://code.google.com/p/misiek/`

Please don't hesitate to contact us with any comments and suggestion or if you are interested in co-developing this software.

1.2 About the authors

This plugin was implemented by Michał Woźniak. Project idea and guidance came from Janusz Dutkowski and Jerzy Tiuryn. Affiliation: Institute of Informatics, University of Warsaw

E-mails: `m.wozniak, januszd, tiuryn@mimuw.edu.pl`

2 Manual

2.1 Requirements

- Cytoscape 2.6.* or 2.7.0
- Java 1.6

Plugin was tested on Windows, Linux and Mac platforms.

2.2 Installation

Download plugin jar file *NetworkEvolution.jar* from our project website:

`http://bioputer.mimuw.edu.pl/modevo`

Just copy downloaded file to *plugins* folder in Cytoscape home directory.

2.3 Examples

Examples of usage are available on the plugin homepage in manual section:
`http://bioputer.mimuw.edu.pl/modevo/manual.html`

3 Input file formats

3.1 Species tree

The species tree should be provided in standard Newick representation: ([children]*)parent.

Example

((dm,ce)Bilateria,sc)Fungi

3.2 Protein family trees

Protein family trees should be provided in an enriched recursive format that is similar to Newick, but which puts siblings from the same species in brackets and adds species information after the closing bracket. This extra information is necessary because we have to distinguish the species from which the nodes are coming from. Each line in the file should describe one protein family tree as follows:

```
([nodeA1, nodeA2, ...]speciesA, [nodeB1, nodeB2, ...]speciesB)familyName;
```

Where each node has recursive format:

```
node := proteinName
```
if node is a leaf, or:

```
node := ([nodeC1, nodeC2, ...]speciesC, [nodeD1, nodeD2, ...]speciesD)nodeName.
```

Example

```
([([ce_13562]ce,[dm_178]dm)bi_i1318]Bilateria,[sc_5512,sc_1833,sc_3895,sc_-
4997]sc)root_222;
```

3.3 Interaction data

Interactions for each of the input species should be loaded in the SIF format. Each line should provide the following three fields for one particular interaction:

```
proteinLabel1 probability proteinLabel2
```

Where the probability is a real number ranging from 0 to 1.

In general the probability value is used to filter out less confident interactions to improve the clarity of the network and efficiency of the analysis. A threshold for the probability value is determined by the user before loading the data. A lower threshold can usually be selected for the root network because it is small and has less edges (one node summarizes an entire family of proteins) so it can be displayed without compromising performance. For larger networks a higher threshold should be used.

Example

```
fu_i4 0.4724 fu_i4
fu_i4 0.1622 fu_i11
fu_i11 0.0654 fu_i11
fu_i11 0.0801 fu_i12
fu_i12 0.0803 fu_i12
```

3.4 PPI experimental data

There is a significant conceptual difference between the experimental data and interaction data (discussed in the previous section). For the interaction data we assume that in each species two proteins can be connected by an interaction and this happens with a certain probability (confidence level). In case of the experimental data, there may be multiple cases in which an interaction was observed (multiple publications). An interaction (from the interaction data) could also not be observed in any particular experiment for a given species, but still have a non-zero probability based on experimental evidence in other species. Providing experimental interactions is optional.

Experimentally identified PPI should be loaded in TAB delimited format. Each line should provide the following four fields for one particular interaction:

```
proteinName1 proteinName2 speciesName experimentID
```

Example

YLR363C YMR080C sc 7883168
YIL061C YMR080C sc 7883168
YMR190C YBR114W sc 9253198
YBR114W YMR190C sc 9253198
YBR217W YPL149W sc 9759731
YPL149W YBR217W sc 9759731
FBgn0004837 FBgn0004647 dm 8749394
FBgn0004647 FBgn0004837 dm 8749394
FBgn0010417 FBgn0000617 dm 8598927
FBgn0000617 FBgn0010417 dm 8598927
FBgn0035371 FBgn0039055 dm 15710747
FBgn0004922 FBgn0000463 dm 15710747

The interactions from the same experiment will have the same color in the visualizations. The tool-tip interface additionally provides the experiment id (e.g. PUBMED ID) for each interaction.