An Integrated Model for Visualizing Biclusters from Gene Expression Data and PPI Networks (IntegratedViz Manual)

Contents

An Integrated Model for Visualizing Biclusters from Gene Expression Data and PPI Networks

Ahmet Emre Aladağ
Computer Engineering
Kadir Has University
Istanbul, 34083, Turkey
emre.aladag@stu.khas.edu.tr

Cesim Erten
Computer Engineering
Kadir Has University
Istanbul, 34083, Turkey
cesim@khas.edu.tr

Melih Sözdinler
Computer Engineering
Boğaziçi University
Istanbul, 34342, Turkey
melih.sozdinler@boun.edu.tr

ABSTRACT

We provide a model to integrate the visualization of biclusters extracted from gene expresion data and the underlying PPI networks. Such an integration conveys the biologically relevant interconnection between these two structures inferred from biological experiments. We model the reliabilities of the structures using directed graphs with vertex and edge weights. The resulting graphs are drawn using appropriate weighted modifications of the algorithms necessary for the layered drawings of directed graphs. We provide applications of the proposed visualization model on the *S.cerevisiae* dataset.

Categories and Subject Descriptors

H.5 [Information Interfaces and Presentation]: Miscellaneous; H.1 [Information Systems]: Models and Principles

General Terms

Design, Algorithms

Keywords

Bicluster, PPI network, Visualization

1. Introduction

Our tool is aimed to combine the concepts of biclusters and PPI networks. We provide a model to integrate the visualization of biclusters extracted from gene expresion data and the underlying PPI networks. Such an integration conveys the biologically relevant interconnection between these two structures inferred from biological experiments. We model the reliabilities of the structures using directed graphs with vertex and edge weights. The resulting graphs are drawn using appropriate weighted modifications of the algorithms necessary for the layered drawings of directed graphs. We provide applications of the proposed visualization model on the S.cerevisiae dataset.

The code is written in both C\C++ and Python languages. In order to run it you need to have a C compiler and Pyhon interpreter. We aimed to give a visualization of biclusters using some biological data and metrics as well. So we provide a graph based visualizations contain both biological data visualization in peripheral graphs and bicluster relevence in main centered graph.

You can easily update the code giving suitable PPI data and bicluster results. We try to make the code easier to understand and modify. Since it is based on Yeast 2884x17 data, make sure that you change the relevant variables when you port your data into these tools. You will also need a gene-categoty file to give it as an input the the layout. This file should contain gene and its categories. Make sure when you update this file, you will also need to update the capital letters provided in data to the source code. We now use 13 main categories as a char variable

char abbv[] = "EGMPTBFOARDUC";

E - energy production
G - amino acid metabolism
M - other metabolism
P - translation
T - transcription
B - transcriptional control
F - protein fate
O - cellular organization
A - transport and sensing
R - stress and defense
D - genome maintenance
C - cellular fate / organization
U – uncharacterized

You need the specify the file "sources/ppi_sources/yeastgenefunctions".

Finally, you need to specify a gene name conversion file if your bicluster results are not in the form of real gene names.

We still improve the layout due to some operations needen in x-coordinate assignment. The implementation is public and written in python. If you check the latest version from this address http://code.google.com/p/pygml/source/browse/#svn/trunk%3Fstate%3Dclosed you can specify the updated codes by uploading into "/src/python" folder.

2. Installation

You need a gcc C compiler and Python interpreter. You also need to install LEDA C++ library version 5.1 for gcc.

To install LEDA, you need to get tar.gz file. The library is commercial, but you can obtain free versions. See the readme file also.

If you extract LEDA files into a folder then you need to specify library path and other issues.

```
export LEDAROOT=/home/<your user name>/LEDA-5.1-complete-FC4_i386-g++-4.0.2-std
export PATH=$PATH:$LEDAROOT/Manual/cmd
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:$LEDAROOT
```

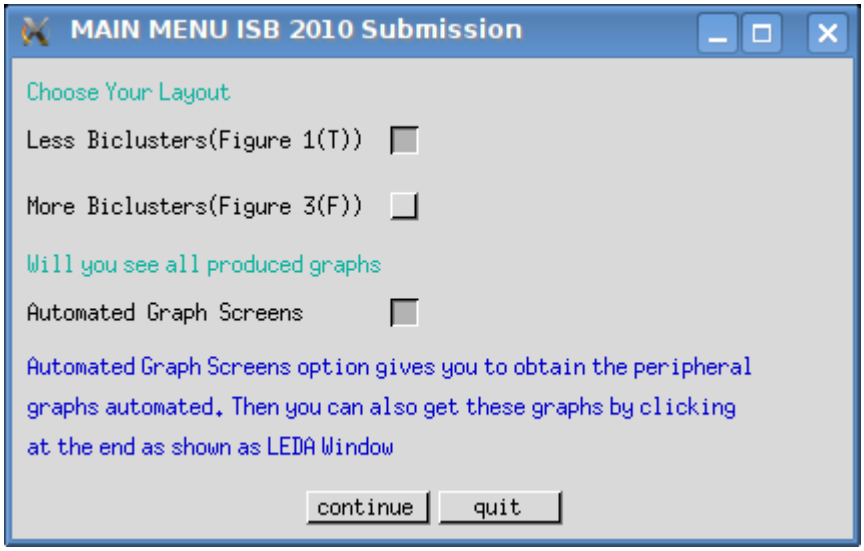
Then simply run gcc with these parameters

```
g++ -I$LEDAROOT/incl -L$LEDAROOT main.cpp -lX11 -lm -lL -lG -lP -lW -o main.out
```

You can run the main.out file as "./main.out" in your linux console.

3. Sample Runs

When you run main.out, you will first see this screen;

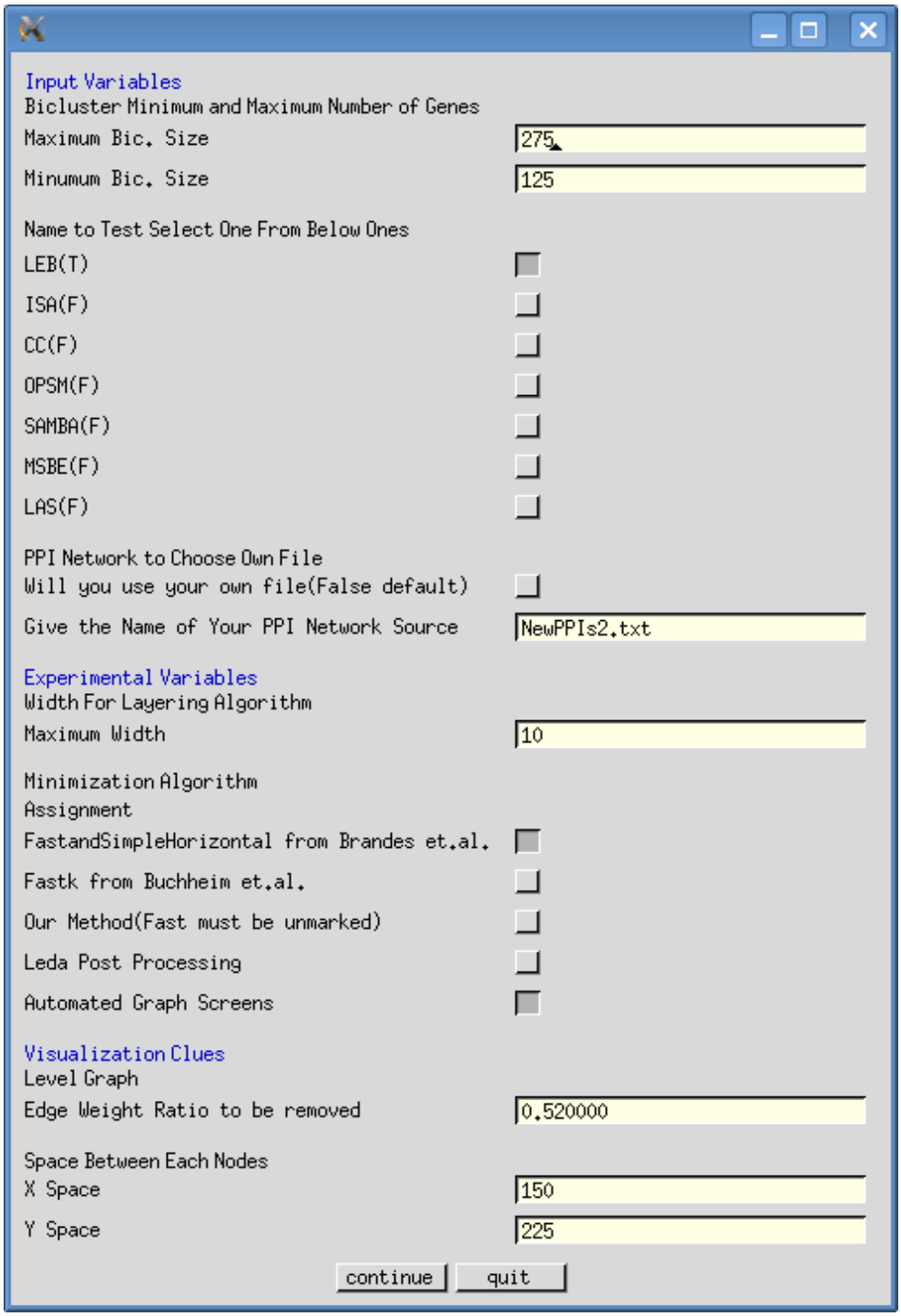


Screen1

This screen will allow you to run Figure 1 and Figure 3 in the paper. You must select one to obtain. If you make multiple selection you will see Figure 3.

"Automated Graph Screens" is an option to see the peripheral graph automatically.

If you press "continue", then you will see the next GUI with more options.

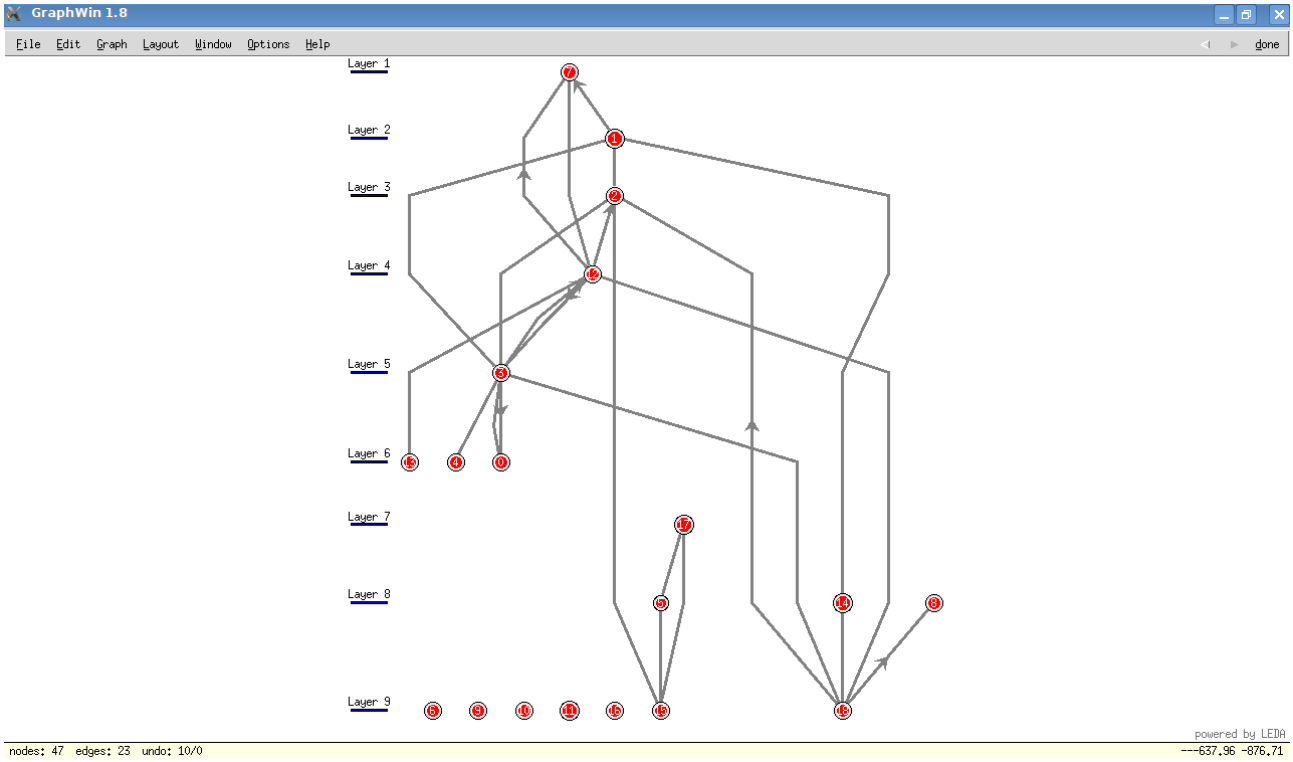


Screen2

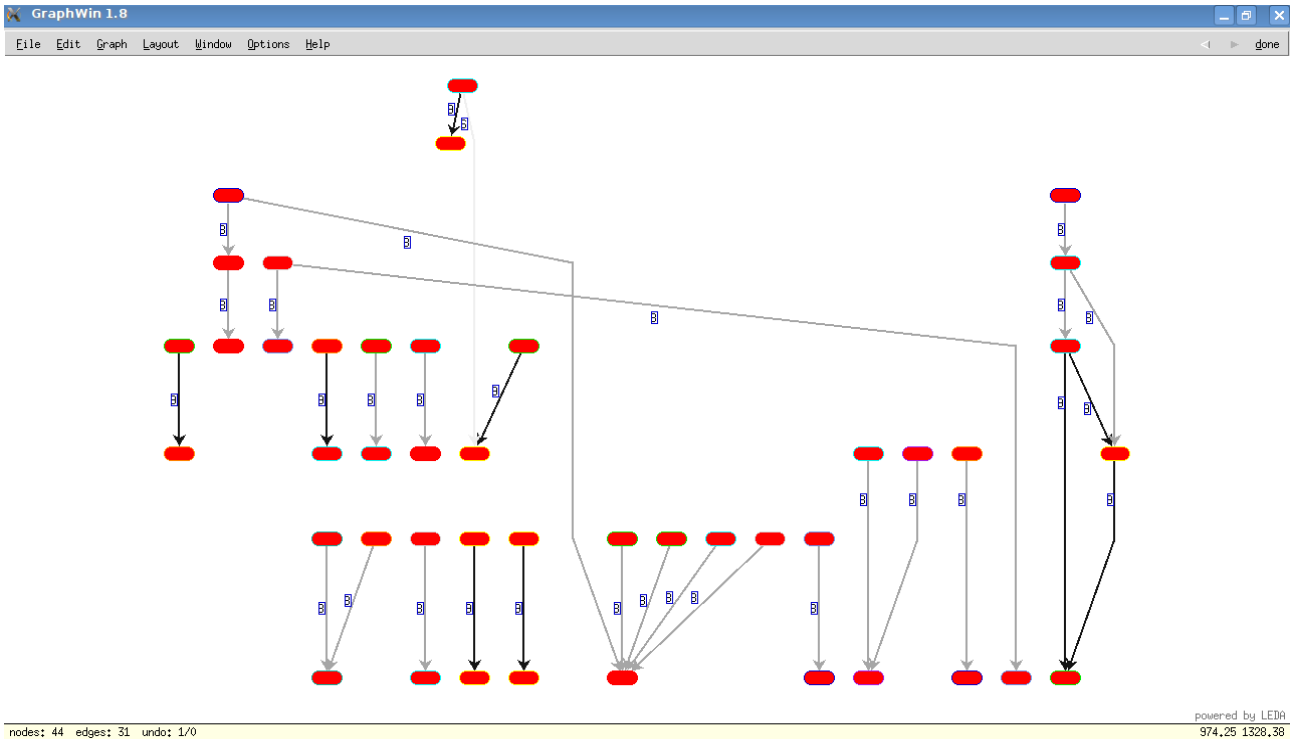
At "Screen2", you can change execution parameters. You can choose the algorithm to import biclusters. You can arrange bicluster sizes and limit the range. You can select different ppi networks related with Yeast organism. You can make filters to relax the layout.

5 - IntegratedViz

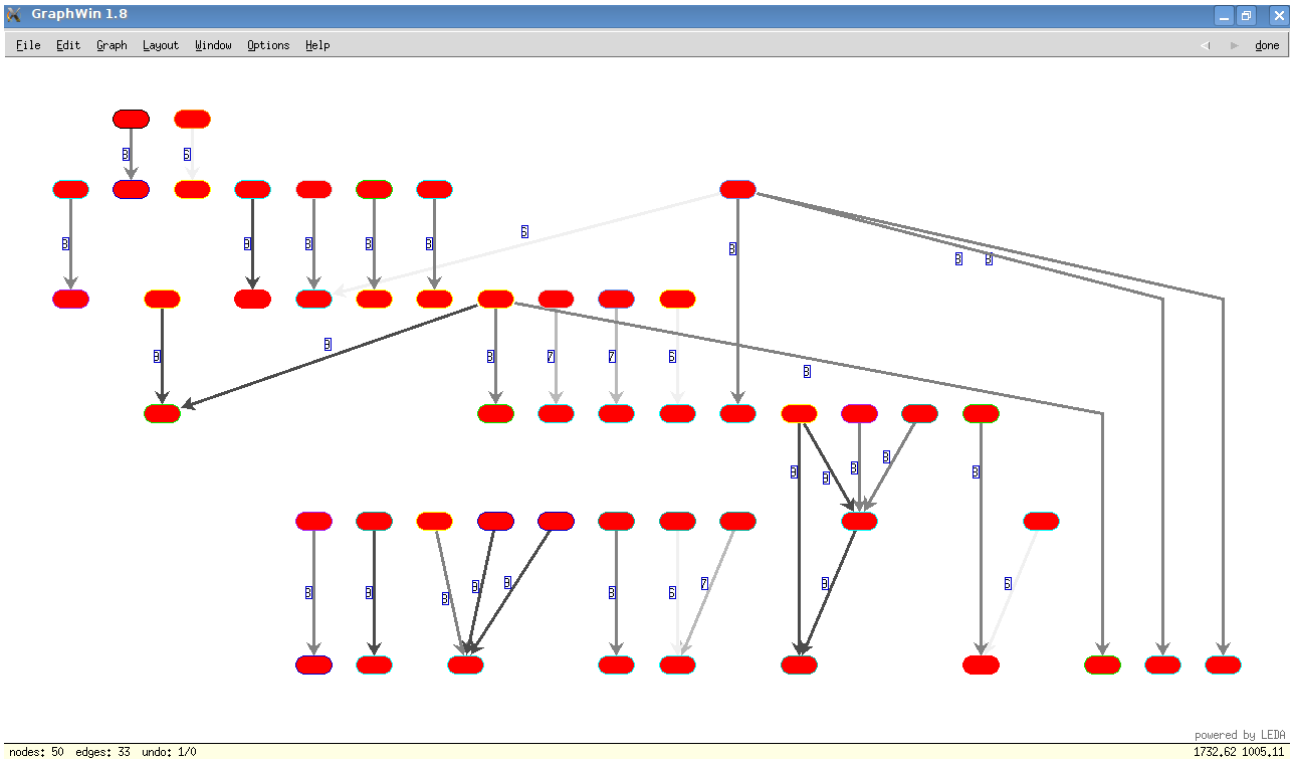
Then the program starts execution. You will see one graphwin screen which shows Layered weighted drawing methodology which is also brand new in the literature. Finally, you can select the color of nodes as you will see middle window and after selection you will obtaing the following figures. Screen 3(figure1 with fast and simple x-coordinate) and Screen 7(figure 3 with fastk x-coordinate) are main central graphs. Others are peripheral graphs for corresponding biclusters of Screen3.



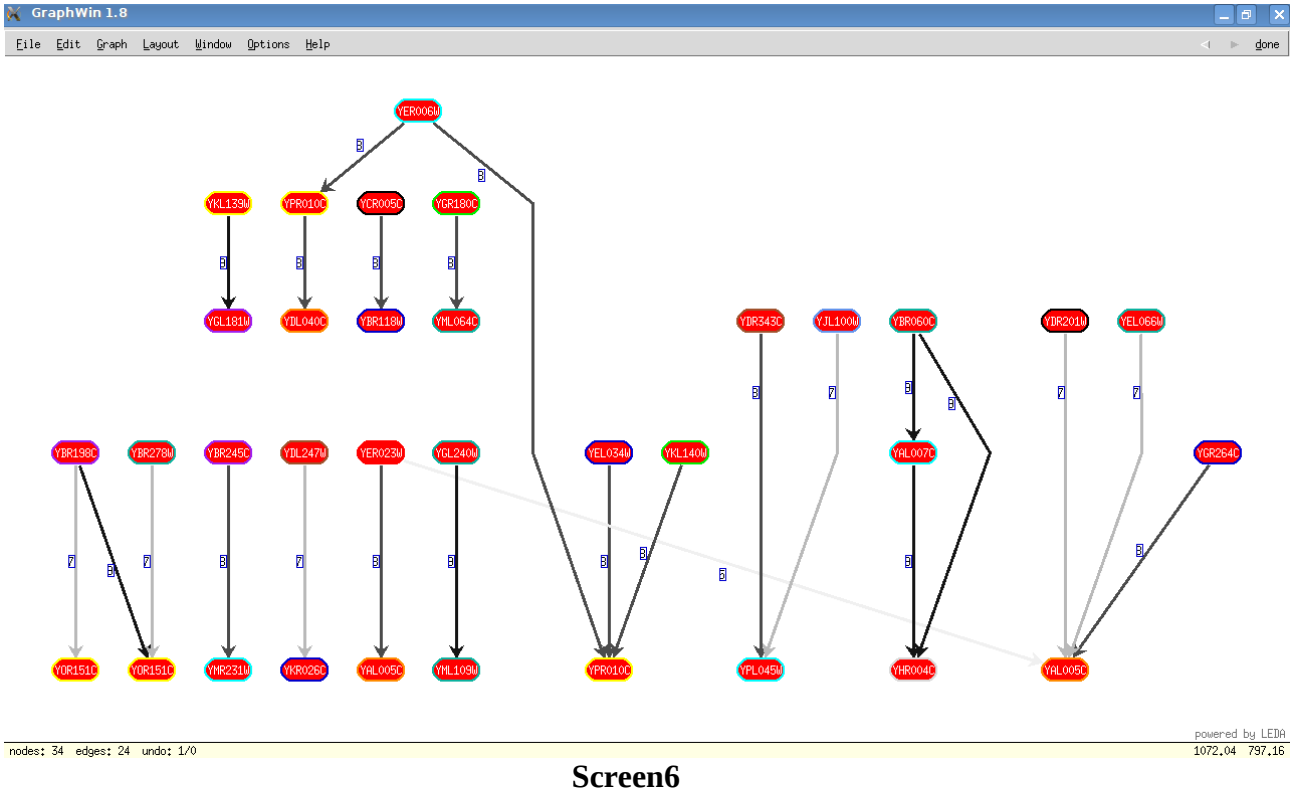
Screen3



Screen4

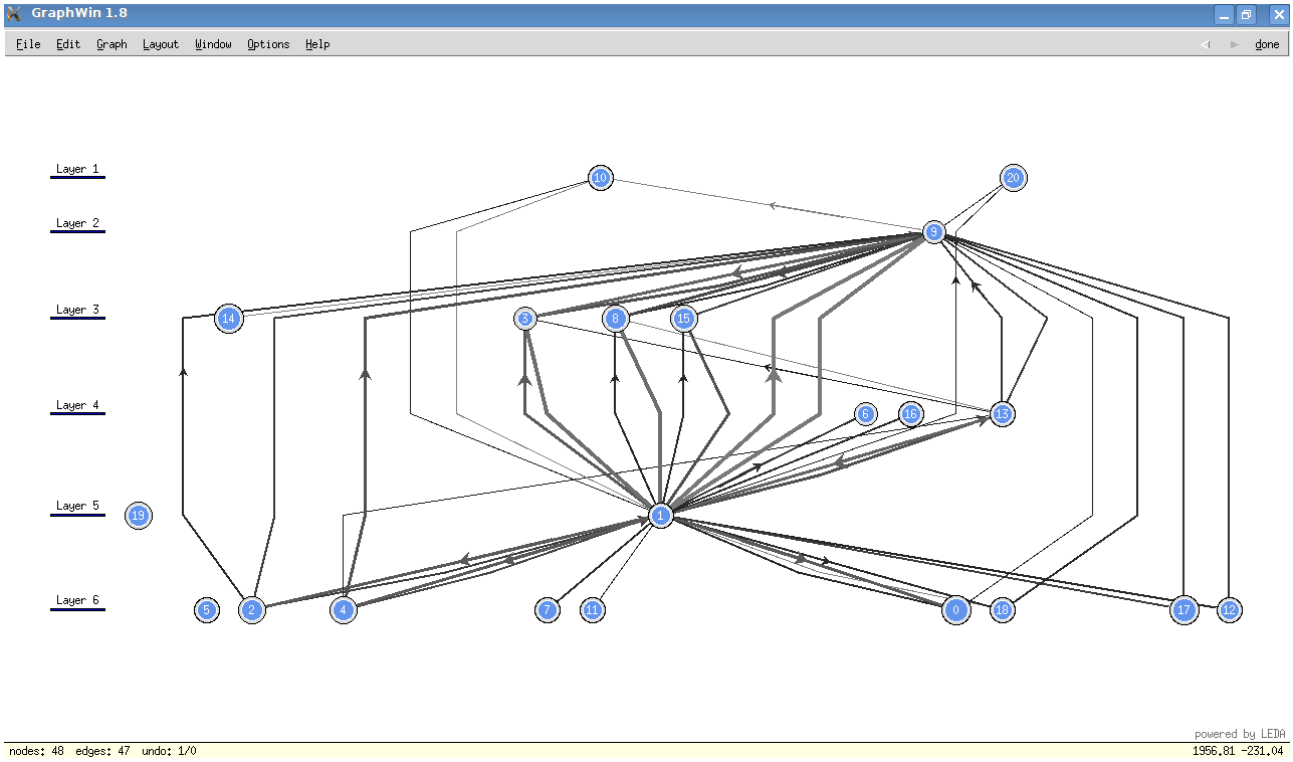


Screen5



Screen6

7 - IntegratedViz



Screen7

4. Citation Issues and License

You can freely use this code and modify under GPL license. Since base library(LEDA) code is commercial you can obtain a license or you can use free version of LEDA. See the website of LEDA(http://www.algorithmic-solutions.info/).

For citation issues, the conference paper at ISB 2010(http://www.isb2010.org/index.html) is accepted and ready for proceeding. You can cite the paper using ACM portal.