TP1 - Introduction to Network Analysis with Cytoscape

Costas Bouyioukos et Anaïs Baudot

First part of the slides

Save your work regularly !!!

Create a protein-protein interaction network for a protein of interest

Import data from PSICQUIC

Opening: from Network database or:
File → Import → Network → Public Databases
Write the name of the protein "HSPB1"→ search

The databases are parsed, and the number of interactions found in each database for the protein of interest is shown. The choice of the database to use will depend on the type of data you want to retrieve: Each database contain different (but overlapping) data, and each have a different philosophy.

Widely used databases include Reactome-FI, Intact and BioGrid, for instance.

Create 2 networks: one from the Intact database, and another one from Reactome-FIs (Merge → No).

Task 1: How many interactions are retrieved from each database for the HSPB1 protein (= what are the sizes of the networks)?

The network created from the reactome-FI database form a connected component: all the nodes are linked to each others, none are isolated. This is not the case in the Intact network, which in addition contains nodes colored differently.

 Task 2: Identify what happened with the Intact network? To this goal, you can use the node annotations given in the Table Panel below the networks

Cytoscape Panels

There 3 different cytoscape panels in the screen: Control panel (with the set of networks, the controls for the visualisation, the apps tabs …), Networks panel (where the network are displayed) and Table panel (with all edge and node annotations).

Manipulate the networks

Visually:
- Try all the buttons zoom in/out, expand on your 2 networks
- Try different algorithms for "Layout"
- Try rotate/align/scale (Layout => Node Layout tools)
- Try rotate/align/scale for a selection of nodes

- In the control panel → Style, modify the default value of node/edge size, colors …

- Choose a satisfying representation for each network (they might be different for the 2 networks), based on all the above tuning

Task 5: Export the network pictures with File → Export as Image.

Adding/filtering data

Filtering data from the Intact Network

The Intact network contains many interactions for the HSPB1 protein, many of which might not be relevant. To remove node/edges, you can directly click and delete them, but you can also create automatic filters.

Create a filter for the network to remove non-human data

Control Panel, tab "Select"
New filter → Column filter → Choose column → Node taxonomy name

This will select on the screen all the nodes that are not annotated as "human" in the Table Panel. These can be removed with Edit → Delete selected nodes and edges.

Task 6: how many nodes/edges remain after this filtering in the Intact network ?

Adding data to the Reactome-FI network

The same process that has been used to create the networks around the protein of interest can be re-applied with the interactors identified in this first step, in order to extend the

network. This leads to the addition of the "interactors of the interactors" and "the interactions between the interactors".

Select all the lines in the table panel of the Reactome-FI network, copy (ctrl-C). Find a smart way to extract only one column containing protein names. Go to File → Import → Network → from public database, and paste the list of proteins.

This will give many more results. Select again "reactome-FIs" and import.

Create a degree/topology filter for this new large reactome-FI network

In the Select tab, create a new Topology Filter, and select the nodes that have a degree of more than 2 (ie, that have more than 2 interactors within distance 1).

Create a new network containing only the selected nodes, and their interactions. For that, you might do File → New Network → From selected nodes, all edges.

Introduction Gene Regulation and Networks

Second part of the slides

Materials

We will base all analyses in two datasets of transcription regulation from RegulonDB, one with the complete TF->target_genes network of interactions and a subset of that which contains only the transcription regulation interactions between TFs (TF->TF network).

Availability

All RegulonDB data are accessible from the following link:
http://regulondb.ccg.unam.mx/menu/download/datasets/index.jsp
We will need the:
1. TF -> Target_genes Regulatory network interactions:
   http://regulondb.ccg.unam.mx/menu/download/datasets/files/network_tf_gene.txt
2. TF - TF Regulatory network interactions:
   http://regulondb.ccg.unam.mx/menu/download/datasets/files/network_tf_tf.txt

Both data files come in a simple text format and contain the following information:
1. A section with comments (start with # and we will ignore)
2. A section with 5 columns containing:
   2.1. TF name
   2.2. Target name
   2.3. Regulatory effect of the TF on target (+ activator, - repressor, +- dual, ? unknown)
   2.4. Evidence that supports the existence of the regulatory interaction

2.5. Classification of the interaction prediction.

The following screenshot shows a part of the top lines of the TF-TF file.

```
#
# Release: 10.5 Date: 09/13/2018
# Genome version: U00096.3, Refseq version; NC_000913.3 GI:556503834
# _____
#
# Columns:
#  (1) Transcription Factor (TF) name
#  (2) TF  regulated by TF
#  (3) Regulatory effect of the TF on the regulated gene (+ activator, - repressor, +- dual, ? unknown)
#  (4) Evidence that supports the existence of the regulatory interaction
#  (5) Evidence Type [Weak, Strong, Confirmed] For details see: http://regulondb.ccg.unam.mx/evidenceclassification
#
AcrR    acrR    -       [AIBSCS, BCE, BPP, GEA, HIBSCS] Strong
AcrR    flhC    -       [GEA, HIBSCS]   Weak
AcrR    flhD    -       [GEA, HIBSCS]   Weak
AcrR    marA    -       [BPP, GEA, HIBSCS]      Strong
AcrR    marR    -       [BPP, GEA, HIBSCS]      Strong
AcrR    soxR    -       [BPP, GEA, HIBSCS]      Strong
AcrR    soxS    -       [BPP, GEA, HIBSCS]      Strong
Ada     ada     +       [BPP, GEA, HIBSCS, SM]  Strong
Ada     ada     -       [BPP, GEA, HIBSCS, SM]  Strong
Ada     aidB    +       [BPP, GEA, HIBSCS]      Strong
AdiY    gadX    +       [BPP]   Strong
AgaR    agaR    -       [BPP, GEA, HIBSCS]      Strong
AidB    aidB    -       [BPP, GEA]      Strong
AlaS    alaS    -       [BPP, GEA]      Strong
AllR    allS    -       [BCE, BPP, GEA, SM]     Strong
AlsR    alsR    -       [GEA]   Weak
```

We will use this very "simple" information to visualise the network and generate some udefull biological information.

Network import and visualisation

Import from file

1. File → Import → Network → from File… (select the TF-TF network file that you have downloaded)
2. Select "Advanced Options", then define comment character, and preview.
3. Select the source and target node (the two first columns), the interaction type (the third column) and the evidence (fifth column) as edge attribute.

Inspect the network and explore different visualisations.

Use different layouts. Download and install the y-files plugin for many sophisticated network layouts algorithms.

Network visualization / manipulation

Select the connected component

Tools → Subnetwork creation → Connected components
(this is the biggest part of the TF-TF network)

Analyse your network

Tools → Network Analyser
Inspect all the different result tabs. What can we say for the diameter, degree distribution and clustering coefficient of the network?

Change nodes, edges sizes

From the respective tabs of the node and edge properties change their respective properties by:
1. Node size → Degree
2. Edge width → Betweenness

What do you observe? Did you gain any biological insight from the current network layout?

Colour edges based on their activity

Based on the information of the third column change the edge colour to represent activating, repressing interactions.

Change edge shape based on interaction evidence

Based on the fifth column change the edge type to represent strong interaction with a solid line and weak with a dashed (or dotted) line.
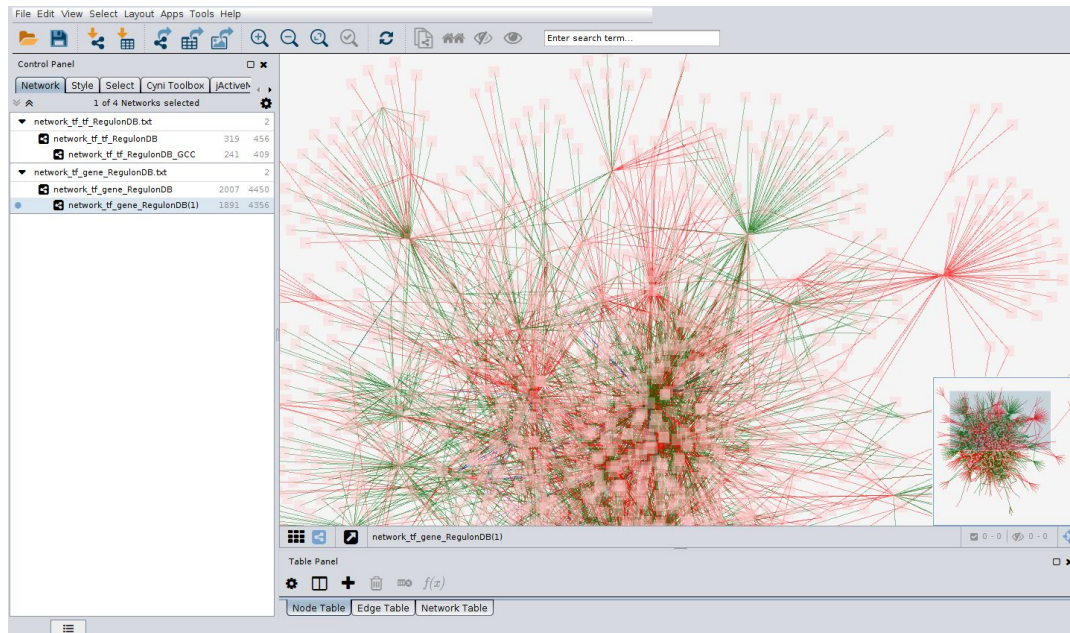
If everything has been done correctly your network will look like this:
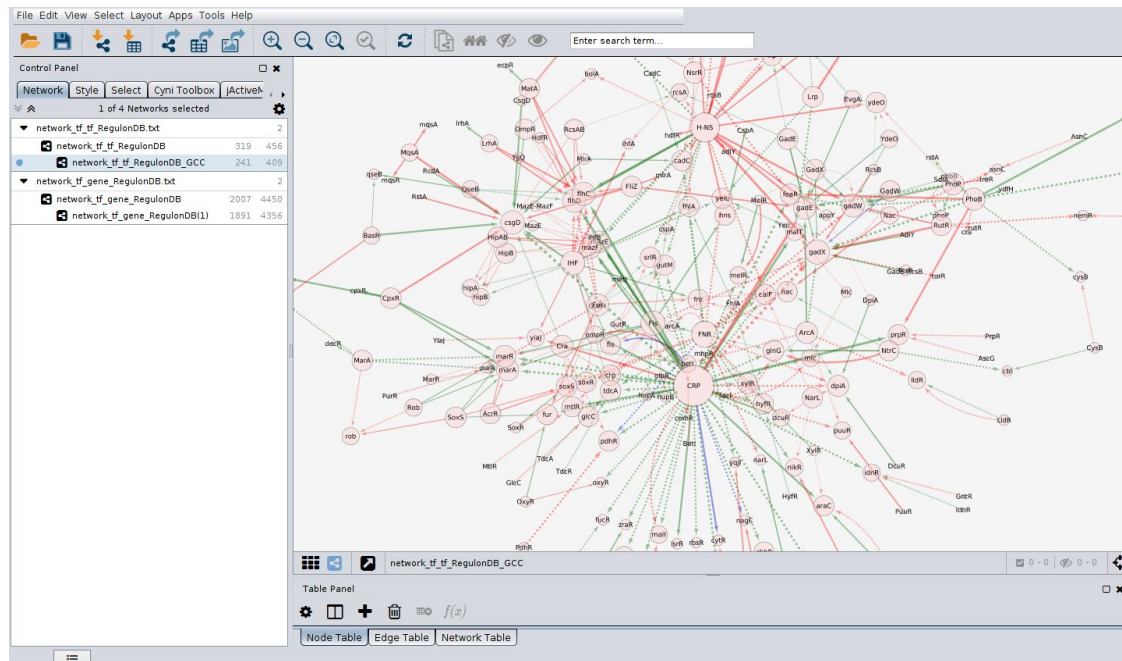
Is there any information useful to biology that is more visible now? Comment on the TFs and on the type and the evidence of the interactions.

Repeat the same tasks for the TF -> Target_genes network.

The full GRN of *E.coli* now should look like this:

Complete regulatory knowledge extraction

Select a TF that you find interesting from your network, go back to RegulonDB website and choose "Regulon list".
You will get a full list of target genes, regulatory sequences and the TF binding motif.

Last part of the slides