Fading Signal Propagation Model

Manual v3

Daniel Rovera,

Team Computational Systems Biology of Cancer
Institut Curie
26, rue d'Ulm
75005 Paris

Contents

This application of Cytoscape 3.x provides functions of visualizing and computing influence spreading to a directed network where edges are activation on inhibition. This application is based on a simple model described in 7).

1   Utils&Analysis

1.1   Preselect Sources and Targets

Preselect in dialogs nodes as sources or as targets. So, these nodes are selected, when opening list boxes. The node attribute PRESELECTED is set to 1 if source, to 2 if target, to 3 if both target and source. If Not preselected it is set to 0. It can be directly imported as node attribute by Cytoscape.

1.2   Select Sub-network from Sources to Targets

This function selects nodes and edges in the directed network between a list of sources and a list of targets. Loops are included in the selection.

1.3   Display SCC in Attribute

Display in node attribute if the node is in a SCC group or not in SCC. The SCC group are numbered (SCC: strong connected components).

1.4   Display Signed Distances

After selecting sources and targets, display the number of paths from sources to targets signed by weights according to the path mode (the parameter "reach" is only necessary for multi path mode).

1.5   List Opened Edges MultiPath Only

Only usable in multi path mode. List back edges which are opened to avoid loop and allow multi path computing.

2   Parameters&Calibration

2.1   Update Weight Attribute

A weight of influence as edge attribute must be affected to every edge. This dialog updates the weight attribute by

- selecting the attribute containing the edge influence,

- affecting its values to activation (weight=+1) and inhibition (weight=-1).
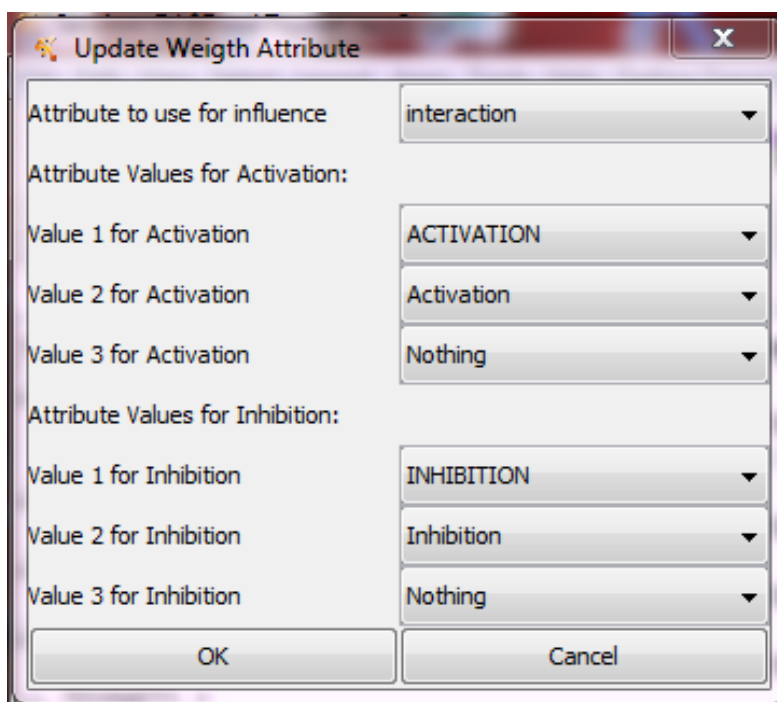
Figure 1: Dialog updating weight attribute from an other attribute

3 possible values for the attribute. Generally, the attribute is "interaction" and the values "activation" or "inhibition". Be careful of lower-case and upper-case (see 1).
The edge attribute "WEIGHT" contains those value which can be input directly by Cytoscape.

2.2   Select Mono or Multi Path Mode

This function selects the type of exploring mode: multi path or mono path.

2.3   Input Reach Parameter

Input the number of paths beyond which the influence is insignificant, less than 5%. It is a floating point number, not necessary integer.

2.4   Input Score Threshold

Input the threshold used to round influence values and, so, to compute kappa.

2.5   Network and Parameter Features

Display in a text box: network, size of network, reach parameter, min, max, mean and standard deviation influence, computed by excluding not connected nodes. It a

recapitulation of parameters and their effect on influence matrix.

3 Numeric Influence

3.1 Display Influence Array For Visualizing

The influence matrix is displayed in a text box which can be copied in the clipboard and paste in a spreadsheet, sources in columns, targets in rows, names in alphabetical order. Parameters and option are in window title. The format is: 3 digits after point for numbers and "nc" = not connected.
Same dialog as "Select Sub-network from Sources to Targets". Preselected sources and targets can be used. For all nodes as sources or targets: select the first node and type shift+control+end.

3.2 Display Influence Array For Computing

Idem but all values are numeric with all possible digits. Less readable but better adapted for computing.

3.3 Display Influence As List

Same result than Display Influence Array As Text For Computing, but as a list between every both nodes.

3.4 Influence by Active Nodes as Attribute

Activity levels of nodes are input in an attribute "ACTIV_IN". The result of the multiplication of influence matrix by activity level input is "ACTIV_OUT" attribute.

3.5 Display Influence Array Between Subnetworks

Display influence array for computing between sub-networks selected in network panel. The influence between sub-networks is the sum of influence of nodes inside sub-networks. The sub-networks must be extracted from a reference network which must be input and which carries the parameters in its attributes. Pieces of incoherence between the sub-networks and the reference network are also displayed in the text box.

4 Graphic Influence

4.1 Display Influence Array as Blue White Red Paved Window

Same computing as "Display Influence Array For ...", but the results are

- as a colored paved window, activated in red, inhibited in blue, light to dark according to the value, not connected in white (see 2).
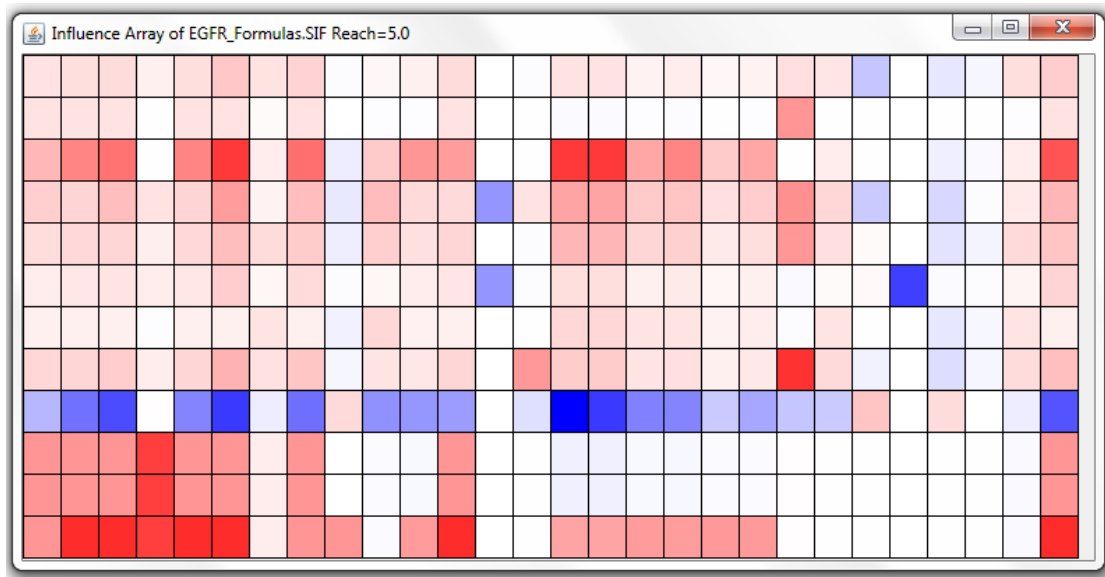
Figure 2: Window paved by the level of influence between species

- when selecting blocs and clicking, influences between nodes in selected blocs are displayed in a text window.

4.2 Display Influence Array as Green Black Red Paved Window

Idem with activated in red, inhibited in green, light to dark according to the value, not connected in black.

5 Reach Area

5.1 Display Influence Reach Area in Array

Computing as "Display Influence Array for Computing" with all weights=1. Useful to appreciate the absolute level of influence by a specie to other species.

5.2 Influence Reach Area as Attribute

Computing from selected nodes with all weights=1. Absolute influence levels are put in attribute "INFLUENCE_AREA_N" where N keeps every successive results. Start nodes and options must be noted manually. Useful to visualize the influence of a group of species.

Set	Input Set Size	Output Aim Size	Sign Score	Active Ok	Inhibit Ok	Kappa
1	13	46	32	19	13	0.385496183
2	13	20	12	9	3	0.157894737
		66	44	28	16	0.320860617

Figure 3: Score of data sets formatted in a spreadsheet

6 Comparing to Measures

6.1 Compute Score of Data Sets

Data must be input as node attributes:

- Activity levels of source nodes as input, attribute "INPUT__SETidentifier" (2 underscores), identifier may be a number or a word,

- Expected activity levels of target nodes as output aim, attribute "OUTPUT_AIMidentifier", identifier matches input with output.

The result is a tabulated text including size of data sets, number of maching output (with threshold) and Cohen's kappa coefficient (see 3 and, about definition see 7). If level <-threshold, node is taken as inhibited. If level >+threshold, node is taken as activated.

6.2 Test Score by Reversing Sign Weight

Test if reversing sign weight of every edge improves kappa, display list of edges sorted by decreasing kappa.

6.3 Test Score by Canceling Weight

Test if canceling sign weight of every edge improves kappa, display list of edges sorted by decreasing kappa. No change in structure network.

7 Description of Fading Signal Propagation Model

7.1 Computing at nodes and along edges

The network for this model is a directed graph where edges are weighted: weight = +1 if activation and weight = -1 if inhibition. Any real value of weight can be used, but the accuracy of this model casts a doubt about this way.

The model is based on fading propagation of the biological signal. The fading effect is simply translated as multiplying edge weights by a global positive factor "fade"(less than 1) along every edge. Fade is translated to a number of paths "reach", which is input, by $fade^{reach} = 5\%$.

Influence on a node is the sum of influences from upstream nodes by incoming edges. This model is the simplest which respect the transitivity by addition of the path lengths as:

$activity(at\ length1, activity(at\ length2, start\ active\ node) =$
$\quad activity(at\ length1 + length2, start\ active\ node)$

More precisely, when going all over the network, the recurrence from every root and every edge (source, target) from the root is:

- $all\_influence = 0$

- $influence(root, root) = 1$

- $influence(root, target) = influence(root, target) +$
  $influence(root, source) * fade * weight(source, target)$

The structure of the network is described by an adjacency list matching every node to lists of adjacent edges. Going over the network can be done by two exploring mode:

1. mono path: exploration by by breadth first search ends when a node already reached is found, every edge edge is counted once ;

2. multi path: loops are opened as far as possible from source by depth first search (back edges), after all paths are explored in order of distances from the source, the exploration stops at 2*reach ;

So, the arrest of exploration at 2*reach keeps multi path process fast (error 0.25%) while coping with forward feed path.

7.2   Comparing to observations

The result is a simple linear model as product of matrices:

- $[Activation\_level\_output] = [Influence\_matrix] * [Activation\_level\_input]$

Activation_level_input can be the states of observed species in input of the network, values being generally:

- +1, if active

- -1, if inactive

- 0 if no observed

Activation_level_ output can be compared to observations of nodes in output of the network.

The computing performance allows to test on every edge weight=0 or weight=-weight to question the relevance of the network.

To compare model results and observations, only the number of matching signs are taken

in a threshold near for computed values. As it exists an alternative, the pertinent indicator is Cohen's coefficient named kappa:

$kappa = (Po - Ph)/(1 - Ph)$

Po is the relative observed agreement,

Ph is the hypothetical probability of chance agreement.

The comparison can concern several sets of observations and, so, it gives a global appreciation of the network.