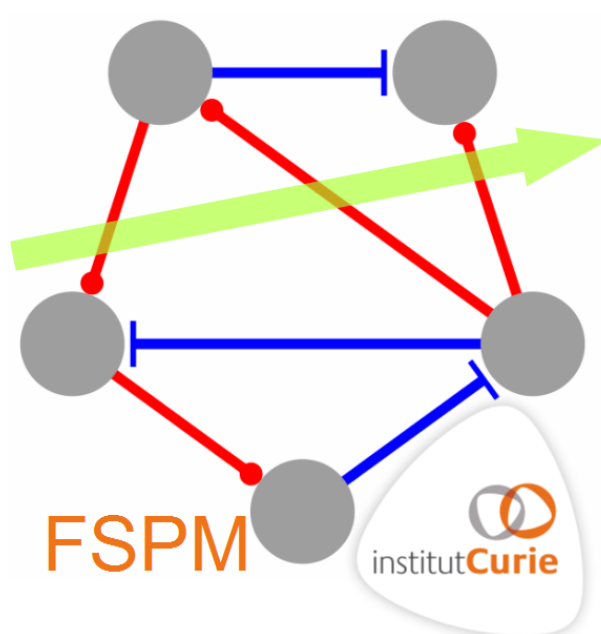


# Fading Signal Propagation Model



## Manual v2

Daniel Rovera - [Daniel.Rovera@curie.fr](mailto:Daniel.Rovera@curie.fr) or @gmail.com

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

# Contents

<b>1</b>	<b>Parameters</b>	<b>3</b>
1.1	Update Weight Attribute . . . . .	3
1.2	Input Reach Parameter . . . . .	3
1.3	Input Score Threshold . . . . .	3
1.4	Preselect Sources and Targets . . . . .	3
<b>2</b>	<b>Structural Analysis</b>	<b>4</b>
2.1	Select Sub-network from Sources to Targets . . . . .	4
2.2	Shortest Path Length Array . . . . .	4
2.3	SCC in Attribute . . . . .	4
<b>3</b>	<b>Model Analysis</b>	<b>4</b>
3.1	Network and Parameter Features . . . . .	4
3.2	List Opened Edges . . . . .	5
3.3	Signed Distance Array . . . . .	5
3.4	Signed Distance List . . . . .	5
3.5	Node List by Path . . . . .	5
3.6	Edge List by Path . . . . .	5
3.7	Weight List by Path . . . . .	5
<b>4</b>	<b>Numeric Influence</b>	<b>5</b>
4.1	Influence Array For Visualizing . . . . .	5
4.2	Influence Array For Computing . . . . .	5
4.3	Influence As List . . . . .	6
4.4	Influence by Active Nodes as Attribute . . . . .	6
4.5	Influence Array Between Subnetworks . . . . .	6
<b>5</b>	<b>Graphic Influence</b>	<b>6</b>
5.1	Influence Array as Blue White Red Paved Window . . . . .	6
5.2	Influence Array as Green Black Red Paved Window . . . . .	6
<b>6</b>	<b>Reach Area</b>	<b>6</b>
6.1	Influence Reach Area in Array . . . . .	6
6.2	Influence Reach Area as Attribute . . . . .	7
<b>7</b>	<b>Comparing to Measures</b>	<b>7</b>
7.1	Compute Score of Data Sets . . . . .	7
7.2	Test Score by Reversing Sign Weight . . . . .	8
7.3	Test Score by Canceling Weight . . . . .	8
7.4	Test Score by Opening Edge . . . . .	8

<b>8</b>	<b>Modification&amp;Simulation</b>	<b>8</b>
8.1	Delete One Node and Reconnect Weighted Edges . . . . .	8
8.2	Effect By Opening Edges with Threshold . . . . .	8
8.3	Effect By Blocking Nodes with Threshold . . . . .	8
<b>9</b>	<b>Check With All Paths</b>	<b>9</b>
9.1	Signed Distance List All Paths . . . . .	9
9.2	Influence List All Paths . . . . .	9
9.3	Influence Array All Paths . . . . .	9
9.4	Node List All Paths . . . . .	9
9.5	Edge List All Paths . . . . .	9
9.6	Weight List All Paths . . . . .	9
<b>10</b>	<b>Interruptible Task</b>	<b>9</b>
<b>11</b>	<b>Description of Fading Signal Propagation Model</b>	<b>9</b>
11.1	Computing at nodes and along edges . . . . .	9
11.2	Comparing to observations . . . . .	10

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 [11]).

## 1 Parameters

### 1.1 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 column attribute containing the edge influence,
- affecting its values to activation (weight=+1) and inhibition (weight=-1).

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 column "WEIGHT" contains the values +1 or -1 which can be modified directly by Cytoscape.

Some functions do not need WEIGHT: Structural Analysis, List Nodes or Edges.

If a column "WEIGHT" already exists and if its type is different from real, this column must be manually deleted.

### 1.2 Input Reach Parameter

Input the number of paths beyond which the influence is not significant, less than 5%. To get any value of fade, reach may be a real, not necessary an integer.

### 1.3 Input Score Threshold

Input the threshold used to round influence values and, so, to compute kappa (for example 0.05).

### 1.4 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.

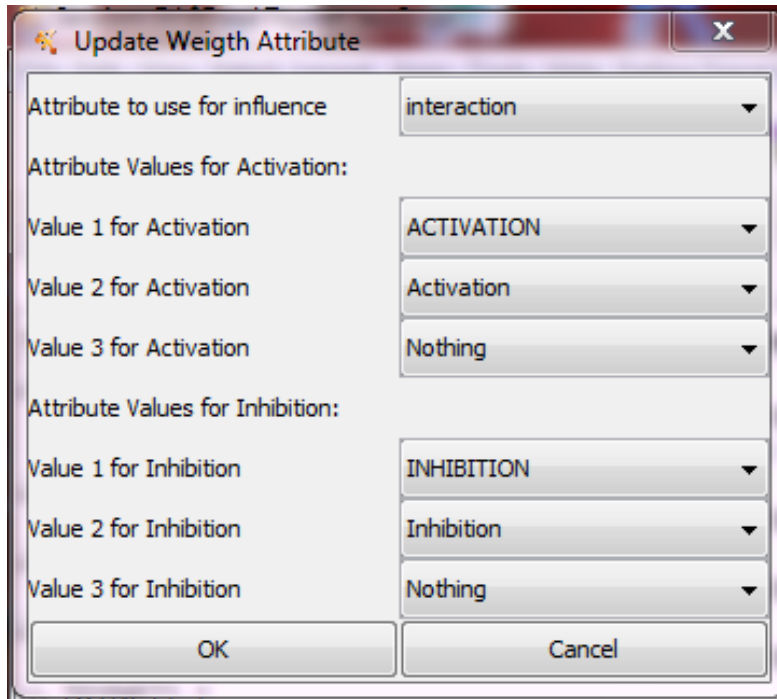


Figure 1: Dialog updating weight attribute from an other attribute

## 2 Structural Analysis

### 2.1 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.

### 2.2 Shortest Path Length Array

Display the shortest path length as an array. "nc" for not connected.

### 2.3 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).

## 3 Model Analysis

### 3.1 Network and Parameter Features

Display 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.

### **3.2 List Opened Edges**

List back edges which are opened to avoid loop and, so, allow computing by recurrence.

### **3.3 Signed Distance Array**

After selecting sources and targets, display as an array the number of paths from sources to targets signed by the product of weights. The parameter "reach" is necessary to limit searching.

### **3.4 Signed Distance List**

Idem but as list.

### **3.5 Node List by Path**

List nodes by path between sources and targets. The network is gone through according to the model 11.

### **3.6 Edge List by Path**

Idem with edges.

### **3.7 Weight List by Path**

Idem with edges displayed as source(weight)target.

## **4 Numeric Influence**

### **4.1 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 rows, targets in column, 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.

### **4.2 Influence Array For Computing**

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

### 4.3 Influence As List

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

### 4.4 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.

### 4.5 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.

## 5 Graphic Influence

### 5.1 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).
- when selecting blocs and clicking, influences between nodes in selected blocs are displayed in a text window.

### 5.2 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.

## 6 Reach Area

### 6.1 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.

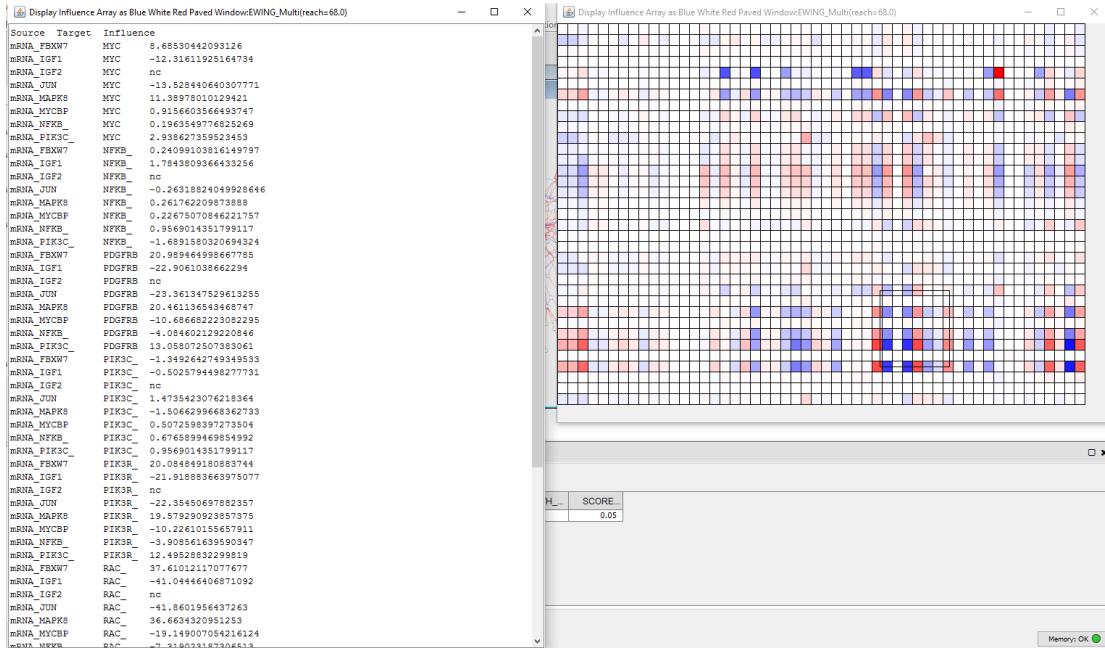


Figure 2: Window paved by the level of influence between species

## 6.2 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.

# 7 Comparing to Measures

## 7.1 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 matching output (with threshold) and Cohen's kappa coefficient (see 3 and, about definition see 11). If level <-threshold, node is taken as inhibited. If level >+threshold, node is taken as activated.



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

## 7.2 Test Score by Reversing Sign Weight

Test if reversing sign weight of every edge improves kappa (no change in network structure). Display edge when a better kappa is found. At end, display list of edges sorted by decreasing kappa.

## 7.3 Test Score by Canceling Weight

Test if canceling sign weight of every edge improves kappa (no change in network structure). Display edge when a better kappa is found. At end, display list of edges sorted by decreasing kappa.

## 7.4 Test Score by Opening Edge

Test if opening every edge improves kappa (change in network structure). Display edge when a better kappa is found. At end, display list of edges sorted by decreasing kappa.

# 8 Modification&Simulation

## 8.1 Delete One Node and Reconnect Weighted Edges

Delete the selected node and reconnect the neighbor nodes . The number of created edges is the product of numbers of in and out edges. The weight of every edge is the product of corresponding deleted edges and interaction is based on sign of product.

## 8.2 Effect By Opening Edges with Threshold

Simulate the opening of every edge and display the effect if the gap is greater than the threshold. Say if there is no change for all couple source, target.

## 8.3 Effect By Blocking Nodes with Threshold

Simulate the blocking of every node as above. The blocked node has no downstream edge.

## 9 Check With All Paths

Functions to evaluate the gap between model and all path. To avoid confusion, the result title ends by -P as path (-M for function using model). These functions need resources, if an exception as Out of Memory occurs, you should try to reduce the scope by limiting the number of sources and targets or by reducing reach.

### 9.1 Signed Distance List All Paths

Idem Signed Distance List for all paths

### 9.2 Influence List All Paths

Idem Influence List for all paths

### 9.3 Influence Array All Paths

Idem Influence Array for computing computed by all paths

### 9.4 Node List All Paths

Idem Node List by Path for all paths

### 9.5 Edge List All Paths

Idem Edge List by Path for all paths

### 9.6 Weight List All Paths

Idem Weight List by Path for all paths

## 10 Interruptible Task

Except graphic influence and results in attribute, all functions are interruptible by click on the cross at the end of the progress bar in task monitor window (see 4). During searching paths in all paths option, the arrest may take time.

## 11 Description of Fading Signal Propagation Model

### 11.1 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

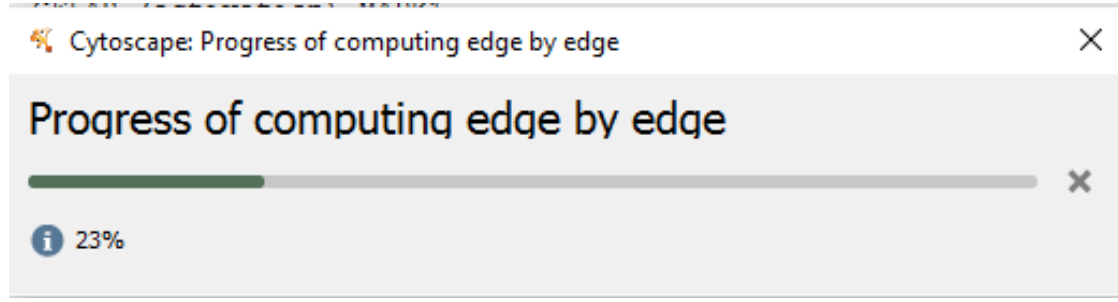


Figure 4: Task monitor window in Effect By Opening Edges with Threshold

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:

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

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 breadth first search ends when a node already reached is found, every 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.

## 11.2 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  $\text{weight}=0$  or  $\text{weight}=-\text{weight}$  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:

$$\text{kappa} = (Po - Ph) / (1 - Ph)$$

$Po$  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.