LitPathExplorer User Manual

Introduction

LitPathExplorer is a visual analytics tool for the joint exploration of pathway networks and biomolecular events extracted from scientific literature. It was designed based on goals which were established as part of several discussions with experts in the respective areas of signalling and metabolic pathway modelling. In particular, we identified four major goals for our tool:

1. Enabling flexible search and exploration of biomolecular pathway networks through the provision of different views of the data as well as various interactive functionalities.
2. We seek to provide a means for making existing evidence in the scientific literature available to support corroboration of information contained in a given pathway model.
3. We aim to facilitate the discovery of potential new interactions that are not yet part of a given model.
4. Within the context of the last two goals, we also quantify confidence in the events, which can provide users of the model with further insight when they are searching for and analysing interactions.
5. We seek to allow the user to become an active participant of the analytical process, where he or she can explore, inspect and revise automatically extracted information, and even potentially facilitate the iterative retraining or recalibration of the system according to this feedback.

The interface of LitPathExplorer is divided into four main areas, as shown in Figure 1. We will explain each of the areas in more detail.

Tutorial video

A four minutes video can be watched to quickly understand how an analyst of a pathway network can benefit from using LitPathExplorer. The video can be found here: http://bit.ly/LitPathExplorer_BI

A - Search area

The search area is the only visible panel at the beginning and it starts with three input text boxes where the user can search for events in a flexible manner. Note that not necessarily all the input text boxes should be filled-in; only one is necessary to perform a search (Fig. 2).

The user can expand the query to form more complex queries by clicking on the green "+" icon. This expands the query to form a Boolean query that gets combined with the one on the left by using and AND or OR operator, as indicated in Fig. 3.

LitPathExplorer

Event

MEK | Role | + -

(A)

10 entities and events found

Network viewer

(B)

ERK
Phosphorylation
Protein_catabolism
Binding
AF-2
MEK
Negative_regulation
Binding
MYC
PPRE

Inspector

Node clicked: Entity (C)

GeneOrProtein:
MEK (PANTHER PATHWAY COMPONENT:P00383)

Events related:

Protein_catabolism (E5163)
Role: agent

Confidence: 0.61
-1   0   1
Polarity: 1.00
Lang. cert.: 0.99
papers: 8
cites: 128
Altmetrics: 21.50
IF: 4.00
Date: Sep 17 2012

Binding (extension_event_1757)
Role: theme

Confidence: 0.58
-1   0   1

Binding (extension_event_1758)
Role: theme

Confidence: 0.58
-1   0   1

Phosphorylation (E2241)
Role: agent

Confidence: 0.61
-1   0   1

Negative_regulation (extension_event_299)
Role: agent

Confidence: 0.57
-1   0   1

Show labels ☑   Discovery ☑   Find in graph | Find
Negatives ☐   Fusion ☑   Layout | Unconstrained ▼
0.56 - 1
Confidence:

Text Analyzer top

Sentences | Word Tree (D)

📄 Article 1 "Zfra affects TNF-mediated cell death by interacting with death domain protein TRADD and negatively regulates the activation of NF-kappaB, JNK1, p53 and WOX1 during stress response." - Published in BMC molecular biology - 2007 Jun 13
Cited by 16 articles - Altmetric score: 3 (63%) - Journal Impact Factor: 2.194

Paper confidence: 1.00
Positive event

In Step 3, **MEK** activates **ERK**, and that Zfra may bind and sequester **ERK** to the cytoplasm.

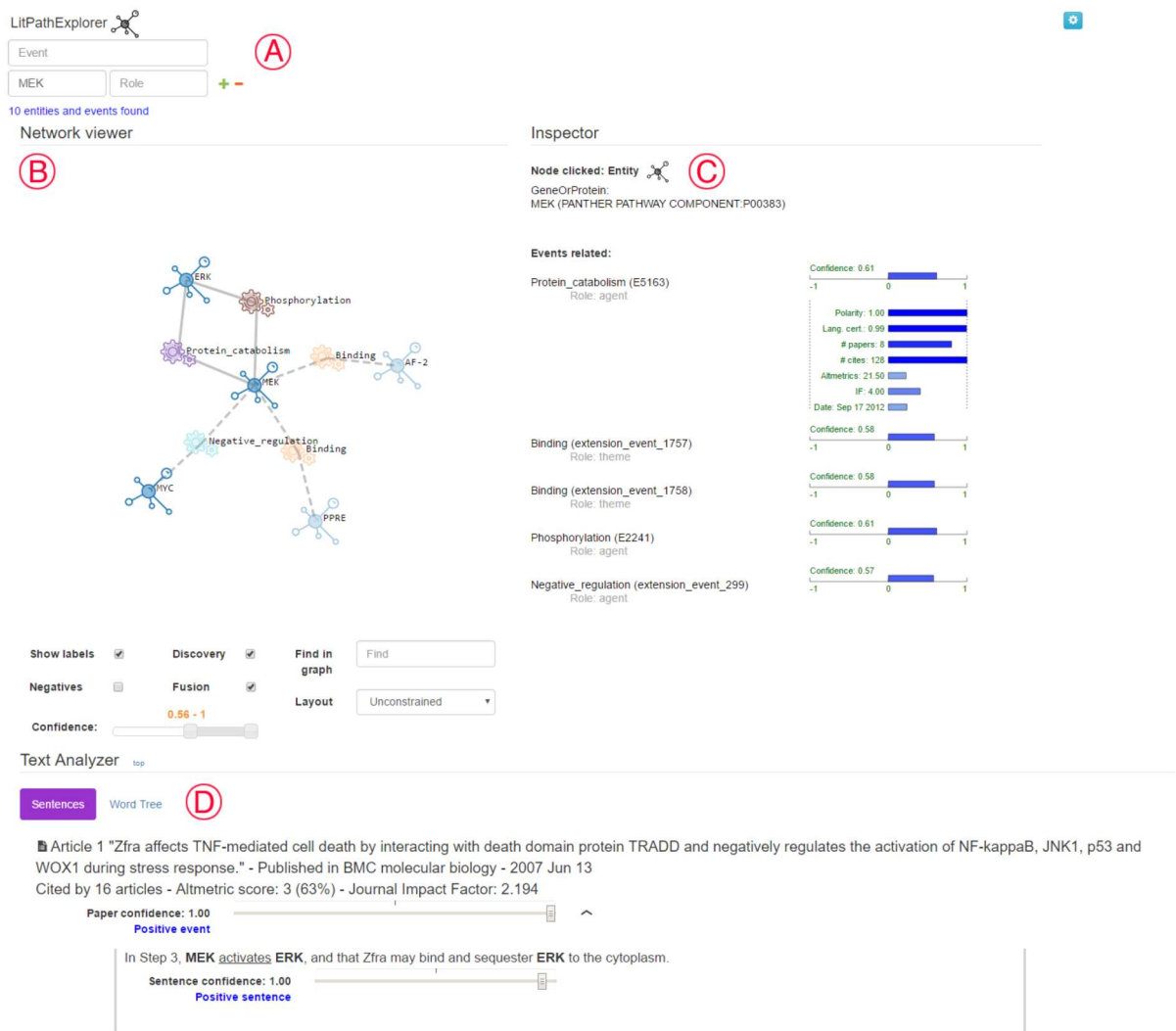Sentence confidence: 1.00
Positive sentence

Fig. 1. Main interface of LitPathExplorer. (A) Search area: allows defining multiple queries to retrieve a subset of the model network. (B) Network viewer: visualizes the model as an interactive node-link graph. (C) Inspector: shows detailed information on the clicked node and its connected events or entities. (D) Text Analyzer: makes available the scientific articles and the specific passages describing a selected biomolecular interaction.

LitPathExplorer

Query for event type → Event

Query for entities that must be in the event → Entity | Role

Role that the entities in the event must play

+ -

Expand to form more complex queries

Fig. 2. Search area

LitPathExplorer

AND ▼ ← Change Boolean operator

Regulation   Event

ERK | Role   Entity | Role

+ - → Reduce components in the query

Fig. 3. Search area after one expansion

B -Network viewer

Our graph-based visualization represents the main component of LitPathExplorer, located on the left side of the interface (Figure 1-B). It visualizes the subset of the model that matched the user's query, based on a nodelink representation: nodes represent entities and events, while links connect event nodes with the entities that take part in the event. Different glyph icons are used to differentiate between entity and event nodes. In turn, different colours are used to represent different types of events available in the model. It is worth noting that links always connect an entity node with an event node.

The size of an event node is proportional to the confidence value computed for the corresponding event. The same analogy applies to the width of the links that connect event nodes, which are proportional to confidence (Fig. 4).
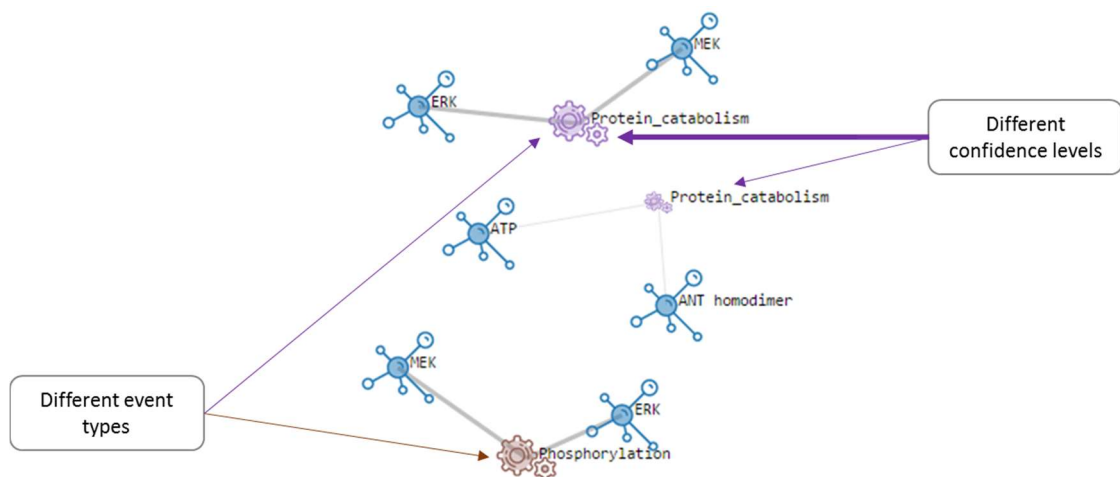
Fig. 4. Events of different type and with different confidence level assigned

Pathway models are enriched with biomolecular events extracted from scientific literature. Events that are not part of the model but involve at least one entity present in the model are referred to as discovered events. Such events are shown in transparent colours (the specific hue dependent upon event type) and linked by dashed lines. Moreover, events whose evidence passages from the literature appear to be contradictory are shown with a glowing yellow border, to enable users to quickly identify them (Fig. 5).
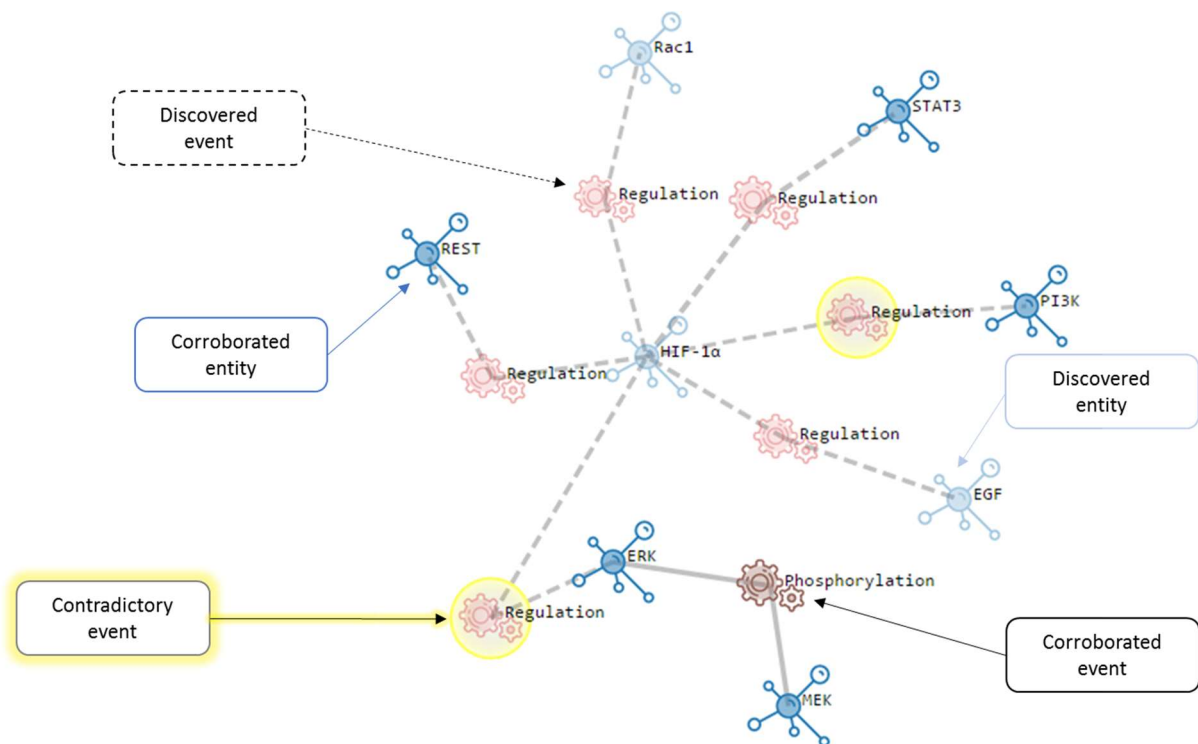
Fig. 5. Discovery mode and contradictory events

Graphs are intuitive representations but have the disadvantage of not being able to scale up well. Dynamic generation of graphs, i.e., drawing them on demand, and interactive capabilities through which users can inspect and modify graph components, have therefore become a means for bypassing this scalability limitation. Some "details on demand" are prompted by different user actions that can be performed on the graph. When hovering the mouse over a node, additional information on the event or entity is shown and their connected nodes are highlighted. Clicking on a node allows the inspection of the event or entity in the inspector panel, while shift-clicking on an entity node expands the graph to show other events that are connected to the clicked node. This functionality complements query-based search and facilitates the visual exploration of the network by querying the network 'on demand'.

Several options for filtering and manipulating the network are available at the bottom of this panel. Maximum and minimum confidence values can be set to filter out events that do not meet the specified range. The layout of the network can be modified although initially it is determined by the force-directed layout algorithm. This panel includes other layouts that organize and sort the event nodes by confidence or aims at separating entity nodes from event nodes. Finally, options for fusing nodes together, as well as for showing/hiding discovered or negated events, are also available.
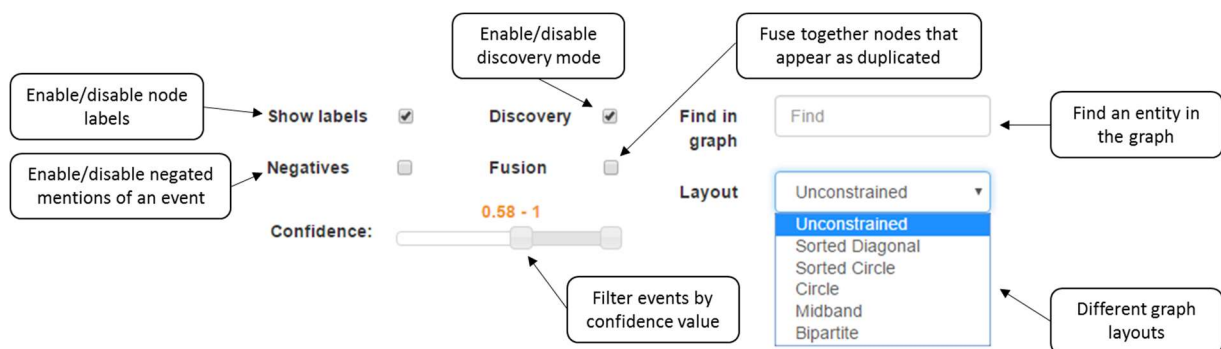
Fig. 6. Further interaction options for interacting with the network

C - Inspector

The inspector panel (Figure 1-C) facilitates closer inspection of entities and events in the model. The nodes selected in the Network viewer can be visualized through this panel, along with a sequential list of all currently visible entities and events that are connected to the selected node. The Inspector is closely coordinated with the Network viewer such that when the label of an event or entity is hovered over, the corresponding node in the network is highlighted, rendering it easy to identify in the other view. A user can utilise this panel to inspect the confidence value computed for every event. Additionally, a bar is displayed to visually encode this value. As domain experts need to understand the reasons behind such confidence assignments, the user can expand the bar plot to show a breakdown of the values that were used to calculate the selected event's confidence value by clicking on the first icon besides the confidence bar.

If further corroboration is desired, the user can inspect evidence found in the literature by clicking on the second icon. A last albeit quite important functionality allows a user to modify an event's current confidence value based on his or her own judgement/knowledge by clicking on the third icon.
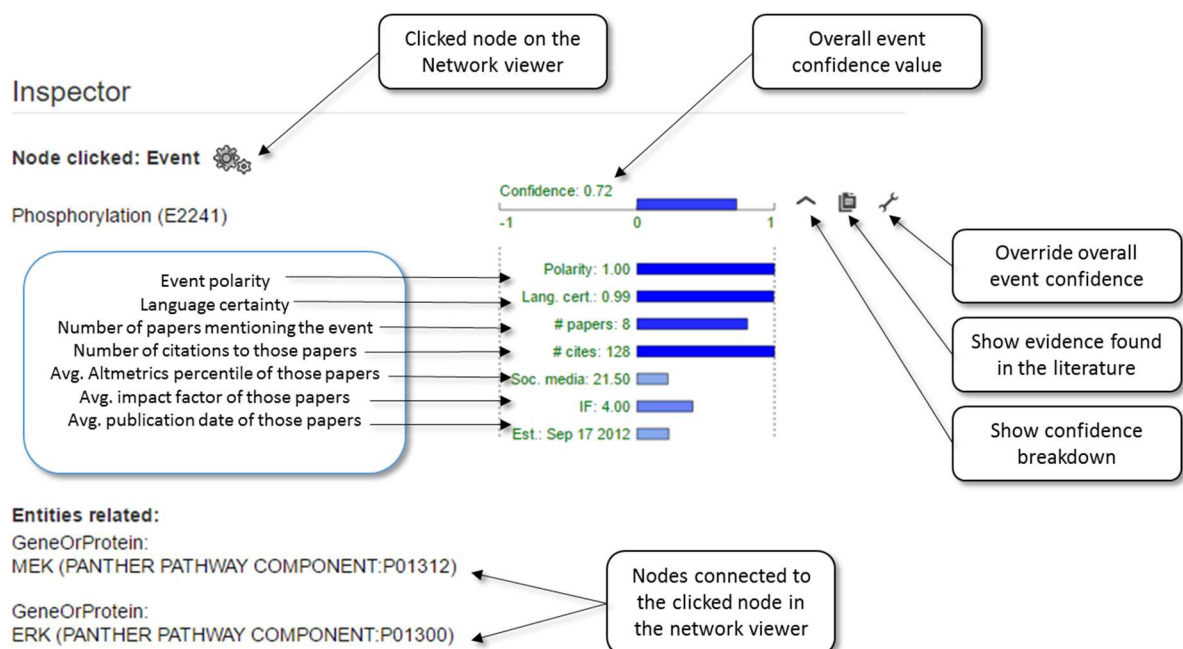

Fig 7. Information about nodes and their connections in the network viewer can be analysed in the Inspector panel. Details on the confidence of each event can be further verified.

If the user decides to override the overall event confidence, a dialog is shown (Fig. 8). If modified, downstream modifications are not allowed (at the paper or the sentence level). Reasons for the override can be indicated to facilitate the history of changes.

In addition, this override can prompt a re-tuning of the confidence values, which can be seen if the neural network-based confidence is enabled (see "Additional Settings" for more information).
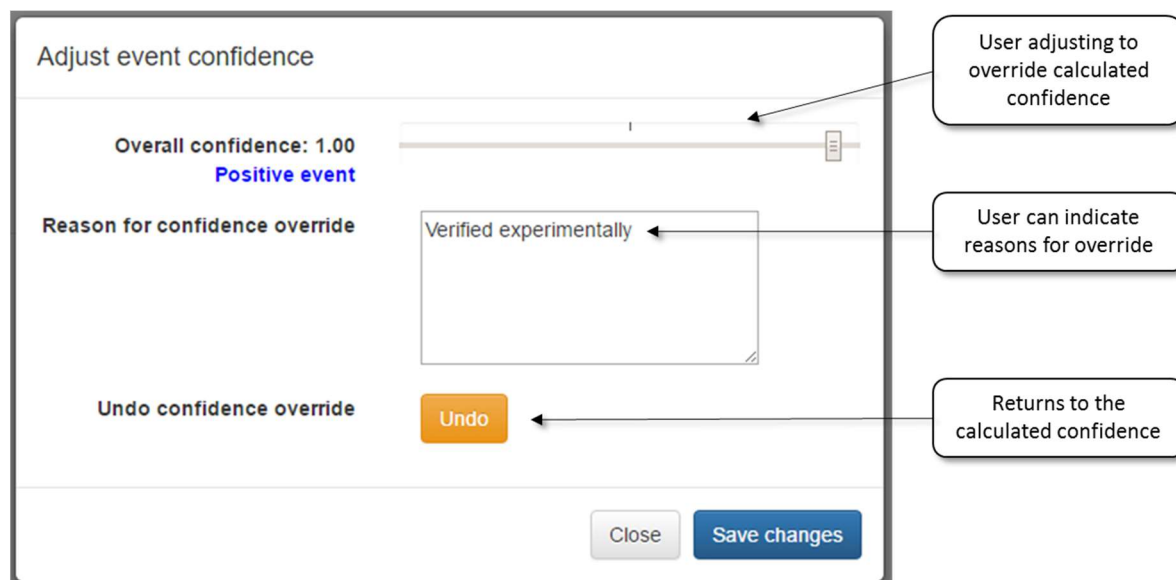

Fig. 8. Options shown when overriding event confidence

In addition each of the entities can be clicked to link to the ontology where this entity is described.

D – Text analyzer

The last panel in the interface is the Text analyzer (Figure 1-D), which visualizes textual evidence found in scientific literature. The textual evidence shown is relative to a specific event which is selected from the Inspector panel.

The information shown in this panel is organized into two different levels (Fig. 9). At the first level are the scientific publications mentioning the specific event, which are shown together with their metadata. At the second level are the sentences or passages containing specific pieces of text describing biomolecular events. Words referring to the entities involved in an event of interest are shown in bold, while trigger words denoting the type of event are underlined. The language-derived confidence value can be inspected and modified, both for each sentence and for an entire paper.

Other interactive features available in this panel include redirection to the PubMed entry of the article and modification of an event's overall language-derived confidence value. This modification initiates several changes, i.e., in the Network viewer, Inspector and Text analyzer panels, which are then propagated to the system's back-end in order to keep track of new user inputs.
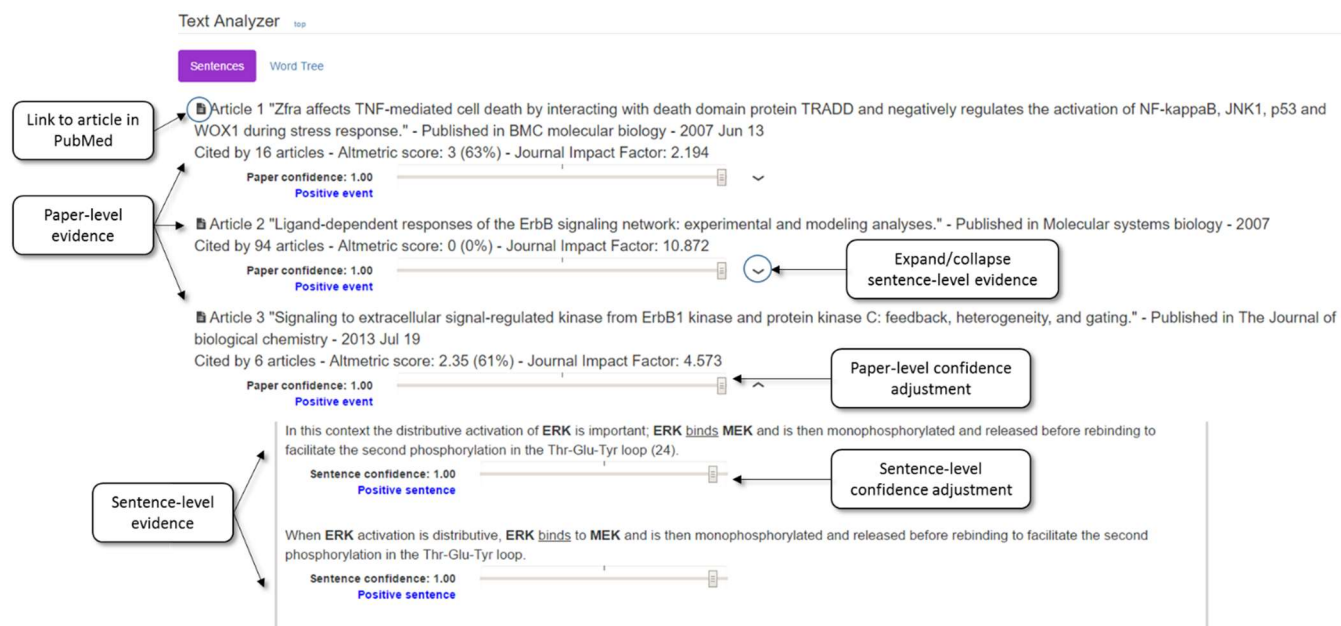
Fig. 9. Text analyser. Showing evidence found in the literature at two levels: paper-level and sentence-level.

Being able to inspect passages helps the user to check the language that was used in reporting a particular biomolecular interaction as well as its surrounding context. However, analysing each sentence in this manner can easily turn into a tedious task when several—and possibly, contradictory—passages from different publications exist. For this reason, we have also implemented a word tree visualization which supports the visual inspection of multiple sentences at the same time (Fig. 10). This visualization builds a suffix-tree (or prefix-tree) based on a given focus word, such that pertinent sentences are arranged in a tree showing all the passages finishing (or starting) with the focus word. In addition, a sentence's position within the tree's vertical arrangement as well as the shade of its font colour, indicate language-derived confidence value. The user can change the focus word as well as switch between visualizing a suffix or prefix tree.
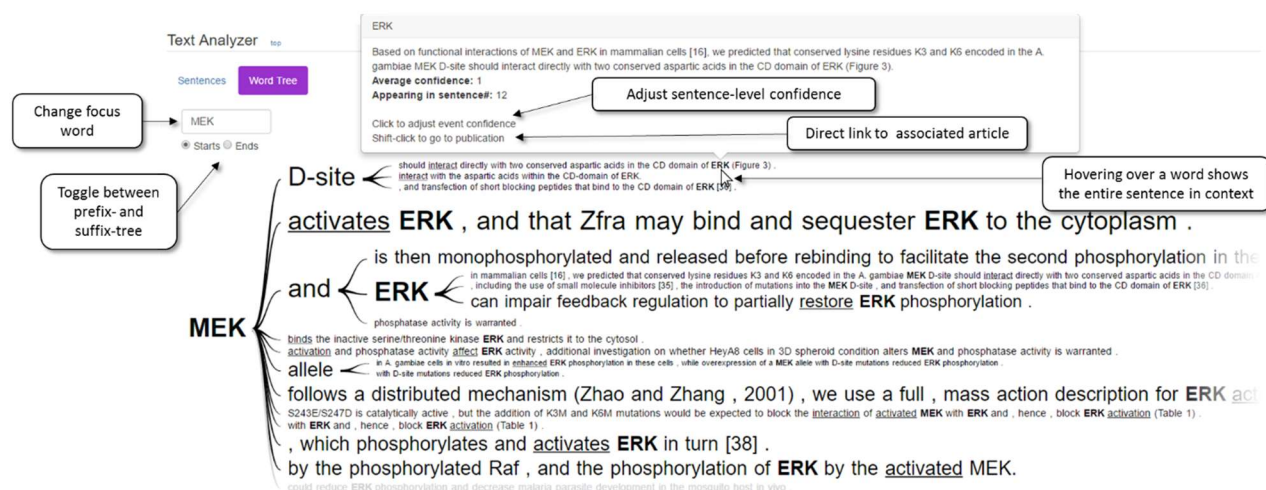

Fig. 10. Word tree visualization. Sentences are shown using a prefix tree. In addition vertical order and shade of colour font denote the associated sentence-level confidence, which can be adjusted upon user interaction.

Additional settings

The user can open additional options by clicking on the gear icon at the top right of the screen. These set of options are shown in Fig 11. The current state of the network can be exported as a CSV file. The destination to where download such CSV file is automatically generated after hitting the "Export" button.

The white list can be used to indicate entities that must be present in the network under analysis. On the contrary, the black list can be used to indicate the entities that cannot be present in the network. Users should indicate one entity per line as indicated in Fig. 11. A blank white list does not impose any restrictions to be visualized.
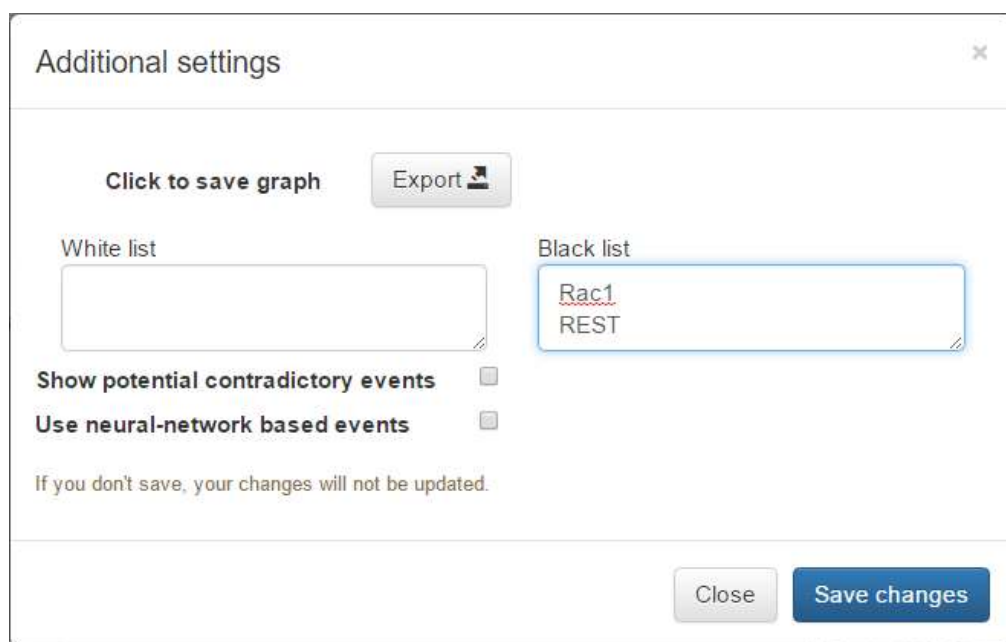

Fig. 11. Additional settings.

Towards the bottom of the dialog there are two checkboxes. The first one is to enable/disable the highlighting on the network viewer of the contradictory events. A contradictory event is defined when a same event has textual evidence supporting it but also negating it. The second checkbox enables/disables the use of a neural network for the computation of the confidences. The main difference that the user will experience is that after overriding an event confidence, the confidence of other weights can also change. The user notes that the confidence weights are being updated by means of a small legend in purple just above the network viewer panel.

Overrides of the event confidences and their corresponding retrained neural network are stored on the server per each different users. This uses a browser cookie to store a unique id per user.

Browser compatibility

The web-based visual search interface is compatible with the following web browsers: Google Chrome, Safari and Opera.

Use of cookies

LitPathExplorer uses cookies to identify previous sessions and recognize previous overrides to event confidences. If cookies are not enabled, LitPathExplorer can still be used, but all the changes applied are nor persistent if the user reloads the tool. No personal information is collected, only the time of the first session along with their changes on the network are saved.

How to cite this work

Axel J. Soto, Chrysoula Zerva, Riza Batista-Navarro and Sophia Ananiadou, "LitPathExplorer: A Visual Analytics Tool for Exploring Literature-Enriched Pathway Models" Under review.