BMC Bioinformatics

RESEARCH ARTICLE   **Open Access**

CrossMark

Diffany: an ontology-driven framework to infer, visualise and analyse differential molecular networks

Sofie Van Landeghem[1,2], Thomas Van Parys[1,2], Marieke Dubois[1,2], Dirk Inzé[1,2] and Yves Van de Peer[1,2,3]*

Abstract

Background: Differential networks have recently been introduced as a powerful way to study the dynamic rewiring capabilities of an interactome in response to changing environmental conditions or stimuli. Currently, such differential networks are generated and visualised using ad hoc methods, and are often limited to the analysis of only one condition-specific response or one interaction type at a time.

Results: In this work, we present a generic, ontology-driven framework to infer, visualise and analyse an arbitrary set of condition-specific responses against one reference network. To this end, we have implemented novel ontology-based algorithms that can process highly heterogeneous networks, accounting for both physical interactions and regulatory associations, symmetric and directed edges, edge weights and negation. We propose this integrative framework as a standardised methodology that allows a unified view on differential networks and promotes comparability between differential network studies. As an illustrative application, we demonstrate its usefulness on a plant abiotic stress study and we experimentally confirmed a predicted regulator.

Availability: Diffany is freely available as open-source java library and Cytoscape plugin from http://bioinformatics. psb.ugent.be/supplementary_data/solan/diffany/.

Keywords: Differential networks, Osmotic stress response, Systems biology

Background

In the early days of Systems Biology, when molecular interaction data was still relatively sparse, all interactions known for a model organism were typically added to a single large interaction network. Such an integrated view would combine data from the proteome, transcriptome and metabolome [1–4]. While such studies certainly proved valuable to gain insights into the general characteristics of molecular networks, they lack the level of detail required to analyse specific response mechanisms of the interactome to changing conditions or stimuli. Consequently, differential networks have been introduced to model the dynamic rewiring of the interactome under specific conditions [5, 6]. Differential networks only depict the set of interactions that changed after the introduction of a stimulus. Most current research in this field has focused on a single interaction type such as expression data [7, 8], genetic interactions [9] or protein complexes [10]. Further, the analysis is usually limited to the comparison of only two networks [11–13]. At the same time, several promising studies have constructed multiple condition-specific networks such as time-course data [14, 15], tissue-specific networks [16, 17] or stress-induced co-expression networks [18]. These studies analyse general network statistics such as connectivity scores or employ machine-learning techniques to identify significantly rewired genes. However, due to the black-box behaviour of the methods and because these studies do not actually generate and visualise differential networks, the resulting prioritised gene lists cannot be easily interpreted by domain experts. By contrast, we believe it to be crucial that researchers can visualise and further explore

*Correspondence: yves.vandepeer@psb.ugent.be
[1]Department of Plant Systems Biology, VIB, Technologiepark 927, 9052 Ghent, Belgium
[2]Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, 9052 Ghent, Belgium
Full list of author information is available at the end of the article

Van Landeghem *et al. BMC Bioinformatics* (2016) 17:18

Page 2 of 12

the rewiring events in their network context. Unfortunately, there is currently no standardised methodology that would allow to integrate heterogeneous condition-specific networks on the one hand, and produce intercomparable differential networks on the other hand.

Here, we introduce a novel ontology-based framework to standardise condition-specific input networks and to allow an arbitrary number of such networks to be used in the inference of a differential network. The network algorithms are designed to cope with a high variety of heterogeneous input data, including physical interactions and regulatory associations, symmetric and directed edges, explicitly negated interactions and edge weights. Depending on the application, these weights may be used to model the strength of an interaction, determined for instance by the expression levels of the interacting genes, or they may represent the probability that an interaction occurs when dealing with computationally inferred networks such as regulatory associations derived from co-expression analysis.

To the best of our knowledge, our integrative framework named 'Diffany' (Differential network analysis tool) is unique in the emerging field of differential network biology, and we hope its open-source release will facilitate and enhance differential network studies. As one such example, we will present how the reanalysis, with Diffany, of a previously published experimental dataset has unveiled a novel candidate regulator for plant responses to mannitol. Experimental validation confirmed that this regulator, HY5, might indeed be involved in the mannitol-responsive network in growing Arabidopsis leaves.

Framework
In this section, we detail the various parts of the Diffany framework (Additional file 1).

Network terminology
To perform a differential network analysis, two types of input data sources are required. First, a *reference network R* models an untreated/unperturbed interactome, serving as the point of reference to compare other networks to. Second, one or more *condition-specific networks* each represent the interactome after a certain treatment, perturbation or stimulus. We denote them as N_i with i between 1 and c, and c the number of distinct conditions that are being compared to the reference state.

Both types of input networks may have edges with a certain *weight* associated to them. Such weights in the networks may be interpreted differently according to the application for which the framework is used. For instance, they may model the strength of physical interactions as determined by expression levels of the interacting genes. In other cases, when dealing with network data inferred through computational methods, such as regulatory associations derived from co-expression data, these weights may instead model the probability/confidence that an interaction really does occur. Whichever the case, the Diffany framework assumes the weights assigned to the edges are sensible and comparable to each other.

The two input sources are used to generate a *differential network D* (Fig. 1) that depicts the rewiring events from the reference state to the perturbed interactome. Further, an inferred *consensus network C* models the interactions that are common to the reference and condition-specific networks, sometimes also called 'housekeeping' interactions. We do not adopt the latter terminology, because while some unchanged interactions may indeed provide information about the cell's standard machinery (i.e. housekeeping functions), others may simply refer to interactions that change under some other condition than the one tested in the experimental setup.
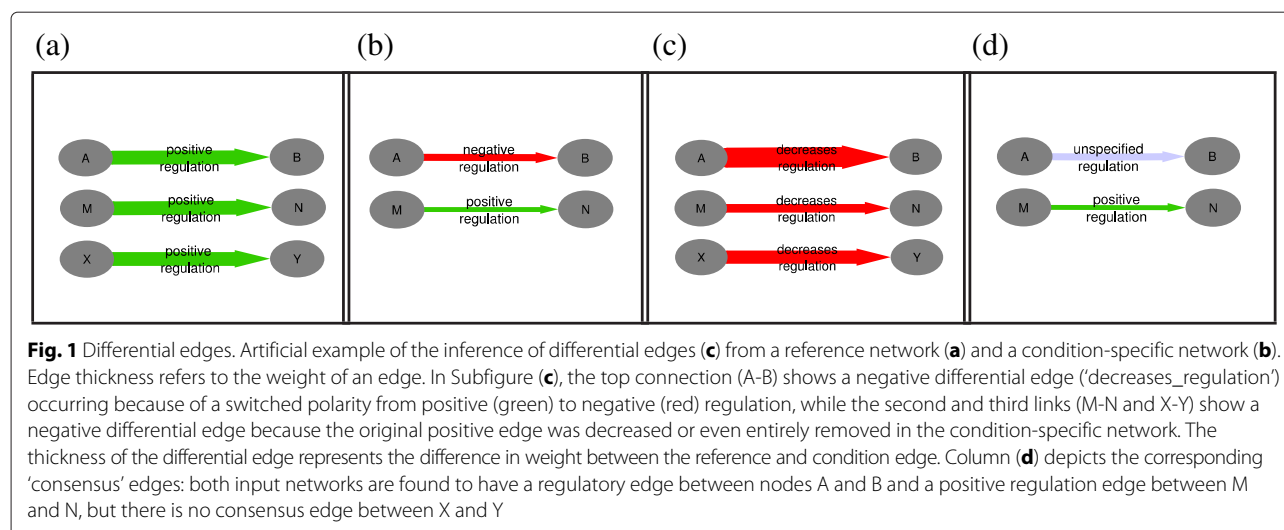
Interaction ontology
The interaction ontology is a crucial component that assigns meaning to heterogeneous input data types. Analogous to the Systems Biology Graphical Notation (SBGN) [19], this structured vocabulory provides a distinction between 'Activity Flow' interactions and 'Process' interactions, modelling regulatory associations and physical interactions separately. However, in contrast to SBGN, these complementary interaction classes can be freely mixed within one network, allowing for a varying level of modelling detail combined into one visualisation.

In the Diffany framework, a default interaction ontology is available, covering genetic interactions, regulatory associations, co-expression, protein-protein interactions, and post-translational modifications (Fig. 2). This ontology was composed specifically to support a wide range of use-cases, and is used throughout this paper. However, the ontology structure itself, as well as the mapping of spelling variants, can be extended or modified based on specific user demands. Additionally, when unknown interaction types are encountered in the input data, they are transparently added as unconnected root categories.

Network inference
The interaction ontology defines the root categories for which consensus and differential edges can be inferred. For the sake of simplification of the formulae in the following, we define $R = N_0$, and we thus have a set $\mathcal{N}$ of $c + 1$ input networks. The union of all nodes in these $c + 1$ input networks is represented by $\mathcal{G}$, and an edge of semantic root category S between two nodes X and Y in an input network N_i as I_{sxyi}. Notice that I_{sxyi} may also refer to a non-existing or 'void' edge when the two nodes X and Y are not connected by any edge of that semantic category S in the network N_i.

Van Landeghem *et al. BMC Bioinformatics* (2016) 17:18

Page 3 of 12

Fig. 1 Differential edges. Artificial example of the inference of differential edges (**c**) from a reference network (**a**) and a condition-specific network (**b**). Edge thickness refers to the weight of an edge. In Subfigure (**c**), the top connection (A-B) shows a negative differential edge ('decreases_regulation') occurring because of a switched polarity from positive (green) to negative (red) regulation, while the second and third links (M-N and X-Y) show a negative differential edge because the original positive edge was decreased or even entirely removed in the condition-specific network. The thickness of the differential edge represents the difference in weight between the reference and condition edge. Column (**d**) depicts the corresponding 'consensus' edges: both input networks are found to have a regulatory edge between nodes A and B and a positive regulation edge between M and N, but there is no consensus edge between X and Y

A differential network is then inferred by considering each possible node pair (X, Y) in $(\mathcal{G} \times \mathcal{G})$ and, for each such pair, constructing the set of input edges $\mathcal{I}_{sxy}$ for each semantic category S. The calculation of differential and consensus edges E from that set of input edges $\mathcal{I}_{sxy}$ involves the determination of the following edge parameters:

- edge negation: $neg(E)$ is a boolean value
- edge symmetry: $symm(E)$ is a boolean value

Fig. 2 Interaction ontology. Default edge ontology structure, with activity flow interaction types on the left, and process types on the right. Root categories are shown with black borders, and have a default symmetry state: directed (→) or symmetrical (-). Because of space constraints, not all PTM (post-translational modification) subclasses are shown
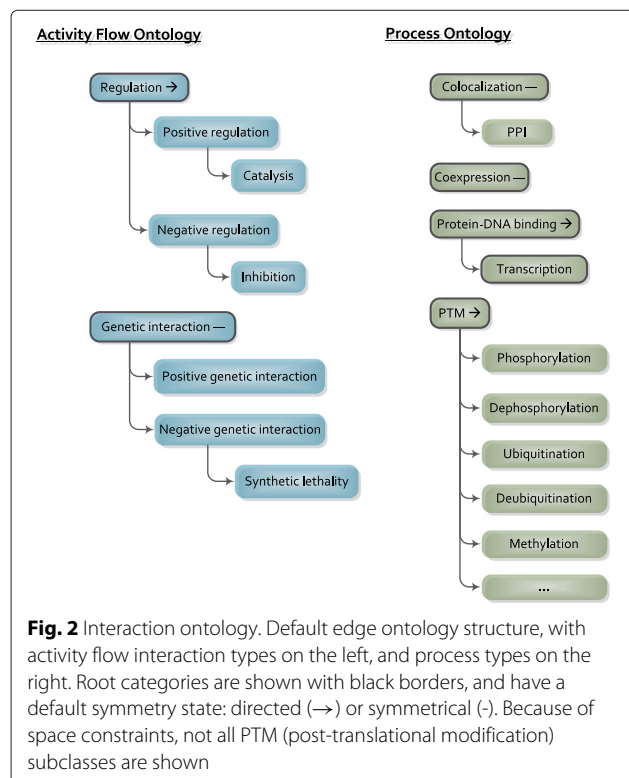
- edge weight: $weight(E)$ is a positive real number
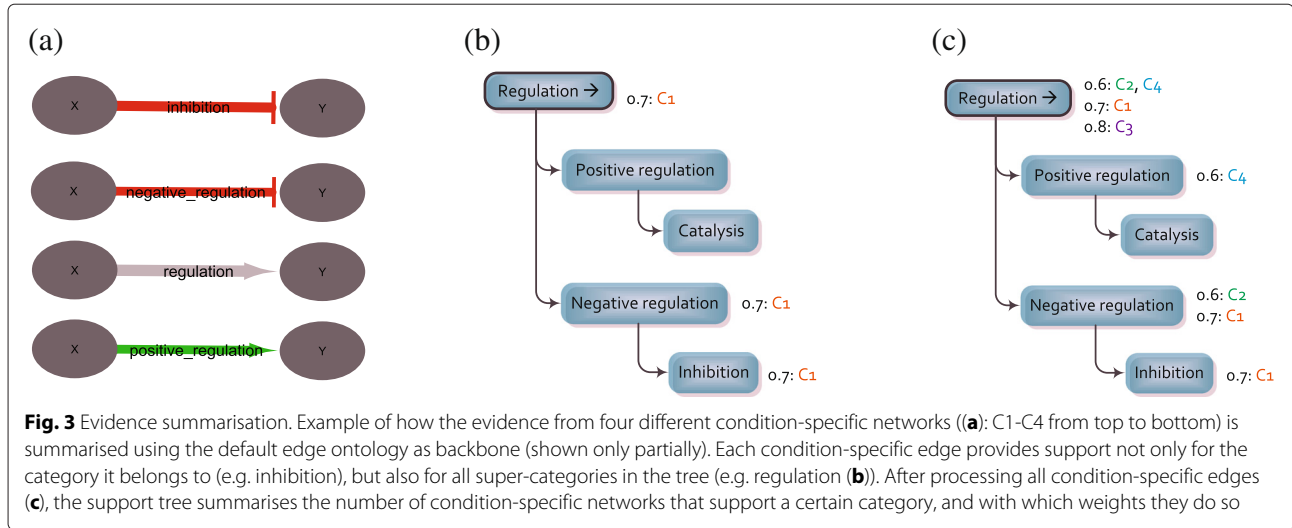- edge type: $type(E)$ is a String value

Differential networks

The hierarchical structure of the interaction ontology forms the backbone for the inference of differential networks. First, all (affirmative) condition-specific edges in $\mathcal{I}_{sxy}$ for a specific category S are processed to construct a support tree (Fig. 3). Such an edge provides support not only for the category it belongs to (e.g. 'inhibition'), but also for all super-categories in the tree (in casu, 'negative regulation' and 'regulation', cf. left tree in Fig. 3). From the support tree that is thus generated, it becomes possible to synthesize the number of condition-specific networks that support a certain category, and by which weights they do so (cf. right tree in Fig. 3).

Negated edges in $\mathcal{I}_{sxy}$ are interpreted as explicit recordings of links that are not present in the interactome, but otherwise do not influence the support tree. A differential edge D_{sxy} is always affirmative (Formula (1)), and is only symmetrical when all input edges in $\mathcal{I}_{sxy}$ are symmetrical (Formula (2)). When only some of the edges in $\mathcal{I}_{sxy}$ are symmetrical while others are directed, the symmetrical ones are unmerged into two opposite directed edges of equal type and weight.

To further determine the type and weight of a differential edge D_{sxy}, the reference edge R_{sxy} is compared to the produced support tree of the condition-specific networks. If the set of values in the support tree (e.g. {0.6, 0.7, 0.8} for 'regulation') contains values both below as well as above the weight of R_{sxy}, no meaningful differential edge D_{sxy} can be deduced, as the response varies in directionality between the different conditions. This is also the case when the edges in $\mathcal{C}_{sxy}$ all appear to be equal to R_{sxy}. Otherwise, when all conditions support a higher weight than the weight of R_{sxy}, the minimal difference to those

Van Landeghem *et al. BMC Bioinformatics* (2016) 17:18

Page 4 of 12

Fig. 3 Evidence summarisation. Example of how the evidence from four different condition-specific networks ((**a**): C1-C4 from top to bottom) is summarised using the default edge ontology as backbone (shown only partially). Each condition-specific edge provides support not only for the category it belongs to (e.g. inhibition), but also for all super-categories in the tree (e.g. regulation (**b**)). After processing all condition-specific edges (**c**), the support tree summarises the number of condition-specific networks that support a certain category, and with which weights they do so

supporting edges determines the increase value shared among all conditions and is thus used as the weight of D_{sxy} (Formula (3)). Similarly, when all conditions support a lower weight, the minimal difference determines the decrease value shared among all conditions. For example, if R_{sxy} would be a regulation edge of weight 0.9, D_{sxy} would be of type decrease_regulation and weight 0.1 according to the support tree of of Fig. 3. If R_{sxy} would have weight 0.4 instead, D_{sxy} would be of type increase_regulation and weight 0.2.

While a Process edge expresses a physical interaction and has no polarity, an Activity flow edge can be determined to have a general 'positive' or 'negative' effect. This means that for an edge in the Activity flow category (e.g. 'positive regulation') also edges of the opposite category can be compared (in casu 'negative regulation'). While in principle edge weights are positive, in this case the weights of the opposite category will be converted to negative values only for calculation purposes. As such, the differential edge between 'negative regulation' of 0.2 (interpreted as −0.2 for calculation purposes) and 'positive regulation' of 0.3 would be of weight 0.5.

$$neg(D_{sxy}) = false \qquad (1)$$

$$symm(D_{sxy}) = \bigwedge_{i=0}^{c} symm(I_{sxyi}) \qquad (2)$$

$$weight(D_{sxy}) = \min_{i=1}^{c} \left(\left| weight(I_{sxy0}) - weight(I_{sxyi}) \right| \right) \qquad (3)$$

Consensus networks
The inference of consensus networks follows a similar procedure. To calculate a consensus edge C_{sxy} from a set of affirmative input edges $\mathcal{I}_{sxy}$, the reference edge R_{sxy} is first added to the support tree in a similar fashion as done previously for the condition-specific edges. The most-specific edge type with highest weight that is supported by all input networks is then chosen to define the consensus edge. In the case when all edges in $\mathcal{I}_{sxy}$ are negated, we construct a similar support tree, but one where the support travels downwards to sub-categories instead of upwards (e.g. 'no regulation' also implies 'no inhibition'). In this case, the least-specific edge type with the highest weight that is supported by all, will represent the consensus edge, which will also be negated (Formula (4)). When $\mathcal{I}_{sxy}$ contains both affirmative and negated edges, no consensus edge will be deduced between nodes X and Y.

As described above, consensus edges are defined by retrieving a weight value that is supported by all input, thus effectively applying a 'minimum' operator to the input weights (Formula (6)). However, it is also possible to apply the maximum operator, which will identify the highest weight that is supported by at least one input network, thus simulating a 'union' operation rather than an 'intersection' between the given input edges. More sophisticated weighting mechanisms will be implemented in the future, depending on the applications in which the framework will be used.

$$neg(C_{sxy}) = \bigwedge_{i=0}^{c} neg(I_{sxyi}) \qquad (4)$$

$$symm(C_{sxy}) = \bigwedge_{i=0}^{c} symm(I_{sxyi}) \qquad (5)$$

$$weight(C_{sxy}) = \min_{i=0}^{c} weight(I_{sxyi}) \qquad (6)$$

Van Landeghem *et al. BMC Bioinformatics* (2016) 17:18

Page 5 of 12

Post-processing

An optional post-processing step is to automatically remove all inferred edges in the differential and/or consensus networks below a user-defined weight threshold. The exact value of this threshold should be chosen based on the input data and the edge weight normalisations of the original resources. For example, the differential weights could be indexed against the null distribution of values expected when the reference and condition-specific networks would represent equal replicates [6].

Fuzzy inference

The differential inference methods as described above can identify a rewiring event that is common to all conditions, as compared to one reference network. However, in some cases it might be beneficial to allow for one or more mismatches. Such a relaxed constraint enables for instance the retrieval of rewiring events that occur in three out of four conditions, thus allowing a more 'fuzzy' or less stringent mode of comparison.

For the calculation of consensus networks, similar relaxed criteria can be applied. In this case, it can be specified whether or not the reference network always needs to 'match' or not. If this is set to 'true', a consensus edge will always need support from the reference network specifically. Otherwise, all input networks are treated as equals.

Implementation

Diffany is implemented in Java 1.6 and the code, released under an open-source license, contains extensive in-line documentation as well as detailed javadoc annotations[a]. JUnit tests ensure proper behaviour of the algorithms also after code refactoring. A GitHub repository provides version control, public issue tracking and a wiki with documentation. For instance, the framework could be extended by adapting more complex statistical scoring strategies [7, 12] into the ontology-based backbone. As this is a non-trivial task, we encourage others to contribute to this effort through the online GitHub repository.

The code base is structured in a modular fashion, with various methods for network cleaning, building and refining the ontology structure, applying custom edge filters, and so on. It is straightforward to extend the available functionality with additional network algorithms or filtering steps. By keeping semantics separate from functionality throughout the code, it becomes straightforward to create a custom ontology for any given project. On top of this core library, we have also implemented a Cytoscape plugin ('app') for the new Cytoscape 3 framework [20], providing an intuitive user interface and allowing straightforward integration with other network inference/analysis tools such as ClueGO [21], BINGO [22] or GeneMANIA [23]. Finally, a commandline interface supports large-scale

bioinformatics studies through the generation of differential networks in straightforward tab-delimited file formats.

Results

By design, the framework presented here can deal with any mixed input networks of negated edges, different edge weights, directed as well as symmetrical edges and a variety of edge types. Herein lays the main strength of our framework that is thus applicable to a wide range of comparative network studies.
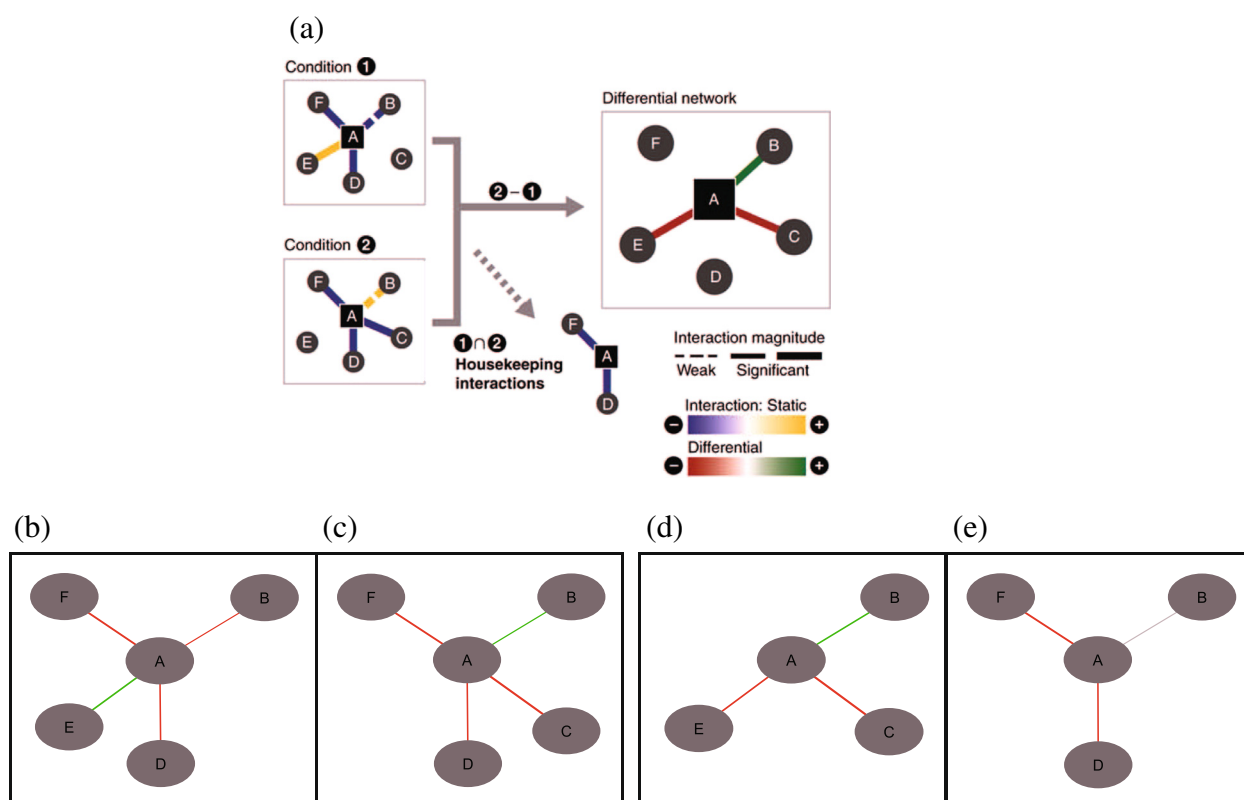
Genetic networks

To evaluate the implementation of our novel framework, we have applied it first to a small, artificial network available in previous literature (Fig. 4). Using the original inference as inspiration (Fig. 4a) to model the input networks (Fig. 4b-c), Diffany produced differential and consensus networks (Fig. 4d-e). Remarkably, compared to the inference of [6], the consensus network generated by Diffany contains one additional edge: the (weak) unspecified genetic interaction (gi) between A and B. Indeed, because our framework is ontology-driven, it can recognise the fact that 'positive gi' and 'negative gi' are both subclasses of the more general category 'genetic interaction'. As a result, there is an edge of type 'unspecified genetic interaction' between nodes A and B in the consensus network.

In cases where such general, unspecified edges without polarity are unwanted, it is trivial to remove them from the network in a post-processing filtering step. However, we believe this additional information can be valuable when combined with the information in the differential networks themselves, as the presence or absence of such a generic consensus edge helps distinguishing between the three different cases as depicted in Fig. 1. Specifically, this generic regulatory edge provides evidence for the fact that both the reference and condition-specific network contain a regulatory edge between nodes A and B, but with opposite polarity, as is the case in the top example in Fig. 1. Given that the differential edge presents an increase in regulation, this means that the reference network contained a negative (down-) regulation, and the condition-specific network a positive (up-) regulation. When instead the consensus edge would not have this general, unspecified edge, as in the case of the bottom example in Fig. 1, this would mean that the condition-specific network simply did not have any link between the two nodes.

Heterogeneous data

The second example presents the application of the Diffany inference tool to heterogeneous input networks, further illustrating the power of the Interaction Ontology. Here, a differential and a consensus network are generated

Van Landeghem *et al. BMC Bioinformatics* (2016) 17:18

Page 6 of 12

Fig. 4 Artificial differential network of genetic interactions. A comparison of Diffany results with a previously published (artificial) differential network involving positive (alleviating) and negative (aggravating) genetic interactions. **a**: The original picture by [6]. The reference network is denoted as 'Condition 1' and the condition-specific network as 'Condition 2'. The differential network is displayed at the right, and the consensus network at the bottom ('Housekeeping interactions'). **b**-**e**: The differential **d** and consensus networks **e** produced by Diffany from the same input data. Because they do not contribute to an enhanced understanding of the molecular rewiring, unconnected nodes are not included in the networks

from reference and condition-specific networks obtained through integrating various interaction and regulation types (Fig. 5). Notice how directionality, different edge types and weights can all be mixed freely in the networks.

Mannitol-stress in plants

To demonstrate the practical utility of our framework, we have used Diffany to reanalyse a previously published experimental dataset measuring mannitol-induced stress responses in the model plant *Arabidopsis thaliana* [24]. In this study, nine-days-old seedlings were transferred to either control medium, or medium supplemented with 25 mM mannitol. At this developmental stage, the third true leaf is very small and its cells are actively proliferating. RNA from these young leaves was extracted at 1.5, 3, 12 and 24 h after transfer. The expression data were processed with robust multichip average (RMA) as implemented in BioConductor [25, 26]. Further, the Limma package [27] was applied to identify differentially expressed (DE) genes at two FDR-corrected P-values: 0.05 and 0.1, giving rise to two sets of DE genes for each time-point (Table 1 and Additional file 2).

Input networks

To determine the set of genes (nodes) relevant to this study, we have first taken all differentially expressed genes across all time-points, using the strict 0.05 FDR threshold. Next, all the PPI neighbours of these genes were extracted from CORNET [28, 29] and added, with the exception of non-DE PPI hubs, as the inclusion of such hubs would extend our networks to irrelevant nodes. Analysis showed that for instance 10 % of all nodes account for 70 % of all PPI edges, and we have removed the bias towards such generic hubs by automatically excluding proteins with at least 10 PPI partners. Note that such hubs will still appear in the networks when they are differentially expressed themselves.

Subsequently, all regulatory neighbours of the extended node set were added, using both the AGRIS TF-target data [30] and the kinase-target relations from PhosPhAt [31]. From the kinase-target relations, hubs with at least 30 partners were excluded, removing mainly MAP kinase phosphatases (MKPs) which are involved in a large number of physiological processes during development and growth [32]. Finally, we also added DE genes from the

Van Landeghem *et al. BMC Bioinformatics* (2016) 17:18

Page 7 of 12
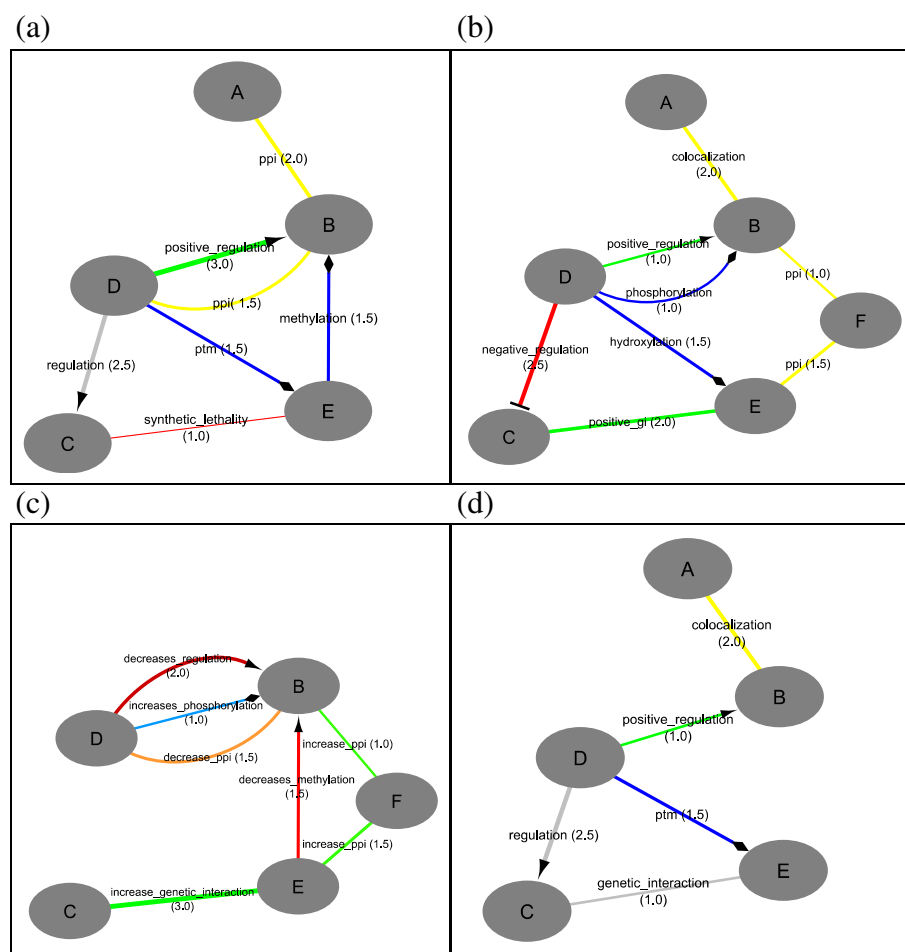
Fig. 5 Artificial differential network of heterogeneous data. More complex calculation of differential (**c**) and consensus (**d**) networks from the reference (**a**) and condition-specific (**b**) networks. Notice how directionality, different edge types and weights can all be mixed freely in the networks

second, less stringent result set (FDR cut-off 0.1), if they could be directly connected to at least one of the genes found up until that point. This approach allows us to explore also those genes that are only slightly above the strict 0.05 FDR cut-off, while reducing noise by excluding those that are not connected to our pathways of interest. In general, this two-step methodology as well as the hub filtering was found to produce more meaningful results. However, both steps are optional and can be removed

Table 1 Number of differentially expressed genes per dataset

Time point	FDR 0.05	FDR 0.1
1.5 h	58	78
3 h	314	456
12 h	435	581
24 h	1500	1913
TOTAL	1638	2155

from the pipeline when using the Diffany library in other studies.

The reference network was then defined by generating all PPI and regulatory edges between the node set as determined in the previous steps. All edges in the reference network were given weight one, a default value used when no overexpression is measured (yet). This resulted in a reference network of 1393 nodes and 2354 non-redundant edges, of which 56 % protein-protein interactions, 24 % TF regulatory interactions and 20 % kinase-target interactions.

Subsequently, each time-specific network was constructed by altering the edge weights according to the expression levels of the corresponding nodes/genes measured at that time point. All interactions with at least one significantly differentially expressed gene as interaction partner is thus down- or upweighted. To define differential expression, the less stringent criterium (0.1 FDR) is used here. For instance, the activation of a non-DE gene by

Van Landeghem *et al. BMC Bioinformatics* (2016) 17:18

Page 8 of 12

a gene that is differentially expressed at that specific time point, would get a weight proportional to the fold change of that differentially expressed activator. By contrast, an edge would be removed (weight zero) when the edge does not fit the expression values at this time point, for instance when an activator is overexpressed but the target is underexpressed. This allows us to remove the interactions that, even though reported in the public interaction data, are probably not occurring in this specific context.

As a final result, the information on differentially expressed genes has now been encoded in the edge weights of the time-specific networks. By comparing them to the generic reference network, the Diffany algorithms will now be able to produce differential and consensus networks which depict the changes in expression values across the time measurements. In the following, we describe these results and provide interpretations that show-case how this type of analysis may lead to novel insights.

Differential network for one condition

With the statistically significant DE values translated into input networks, the differential networks can then be generated by either comparing the reference network to each time-specific network individually, or by comparing all time-specific networks against the reference network simultaneously.

As an example of the first mode of comparison, Fig. 6 depicts the differential network after 1.5 hours, illustrating the rewiring events occurring in this short time frame after the induction of mannitol stress. At this early time point, it is rather unlikely that the expression of the DE genes was affected by subsequent transcriptional cascades. By including transcription factors upstream of the DE genes in the network even if they are not DE themselves, it is possible to identify new putative regulators as compared to previous analysis methods. For example, HY5 and PIL5 might be suitable candidates, as they contain a putative phosphorylation site and are thus likely to be posttranslationally regulated.

To further investigate the possibility that HY5 would be a transcriptional regulator under mannitol stress, we validated the Diffany results by measuring the expression level of the proposed HY5-target genes in the growing leaves of WT and HY5 loss-of-function mutants. These genes, except *ARL*, were all underexpressed in *hy5* mutants as compared to WT, confirming that HY5 is indeed involved in the regulation of the *MYB51*, *EXO*, *RAV2* and *TCH3* expression in growing Arabidopsis leaves (Additional files 3 and 4).

To further explore if HY5 is involved in leaf growth regulation under mannitol stress, phenotypic analysis was performed on *hy5* mutants under both long term and short term mannitol treatment. The *hy5* seedlings were clearly hypersensitive to stress, with decreased leaf size under long term and short term stress, and showed complete bleaching upon long term mannitol stress (Fig. 7, Additional file 4). These biological results demonstrate that HY5, which has been identified with Diffany as a putative regulator of mannitol stress, might indeed be involved in the mannitol-responsive network in growing Arabidopsis leaves.
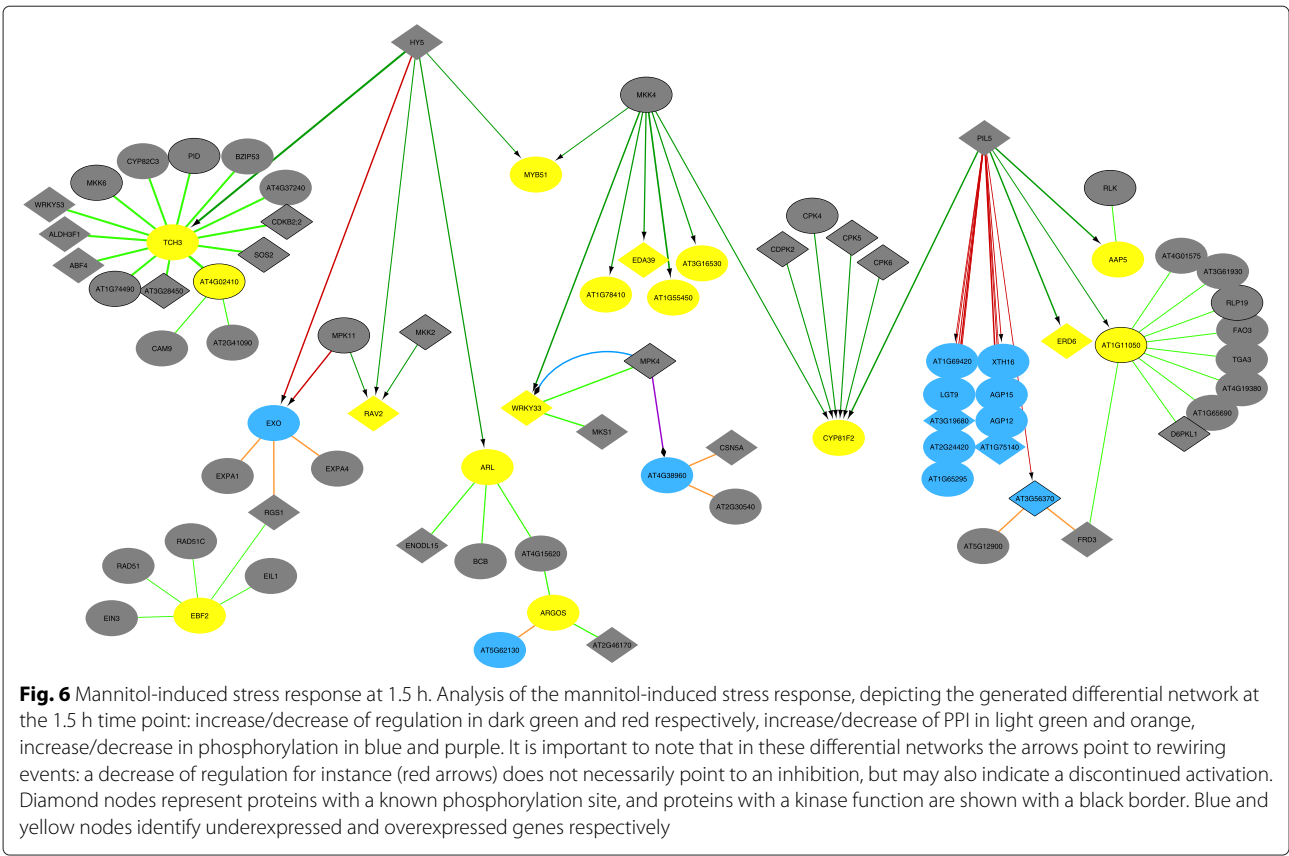
Next to the identification of new putative regulatory links, the differential PPI edges make it possible to understand complex formation under specific conditions. For example, the EBF2 sub-complex presents a nice example of how the induction of one protein is sufficient to increase the activity of a whole complex. The EBF2 is a stress-responsive E3-ligase involved in the posttranslational regulation of the ethylene-responsive factors EIN3 and EIL1 [33, 34]. In this differential network, EBF2 forms a complex with these two targets, which are induced by mannitol as well. However, some of the other members of the SCF-complex, such as CUL1, SKP1, ASK1 and ASK2, are missing from the differential network. As these SCF-complexes are involved in many cellular processes, their specificity being defined by the E3-ligase, we can speculate that the other members of the complex are highly abundant and not specific to mannitol-stress. Their automatic removal from the differential network thus allows the user to focus on the truly interesting genes for this specific stress condition.

Differential network for all conditions

The second mode of comparison allows to simultaneously compare all condition-specific networks to one reference network. In this specific case, such an analysis models the stress-specific, but time-independent response. Fig. 8 shows these rewiring interactions. Strikingly, mainly the overexpressed genes (yellow nodes) remain differentially expressed throughout the time-course experiment, while this is only the case for a few of the underexpressed genes (blue nodes). This implies that in this context, the upregulation of genes is a more stable and long-term process.

For instance, the upregulation of *TCH3* by HY5 is present because *TCH3* is overexpressed at all time points and its upregulation by HY5 may thus play a significant role in the overall stress response. To validate this biologically, the expression level of *TCH3* and other previously mentioned HY5 target genes was measured in WT and *hy5* mutants, 24 h upon transfer to control or mannitol-supplemented medium (Additional file 4). While the induction of *TCH3*, *MYB51* and *ARL* could be clearly observed in WT plants, a more variable but less pronounced upregulation was observed in *hy5* mutants. Thus, HY5 might be involved in the regulation of *TCH3*, *MYB51* and *ARL* under mannitol, although it is probably not the sole regulator of these targets, but instead acts in
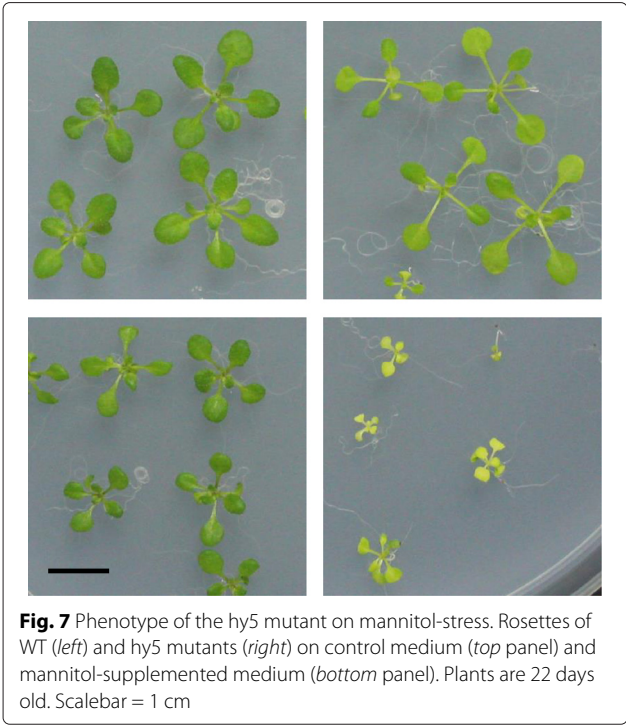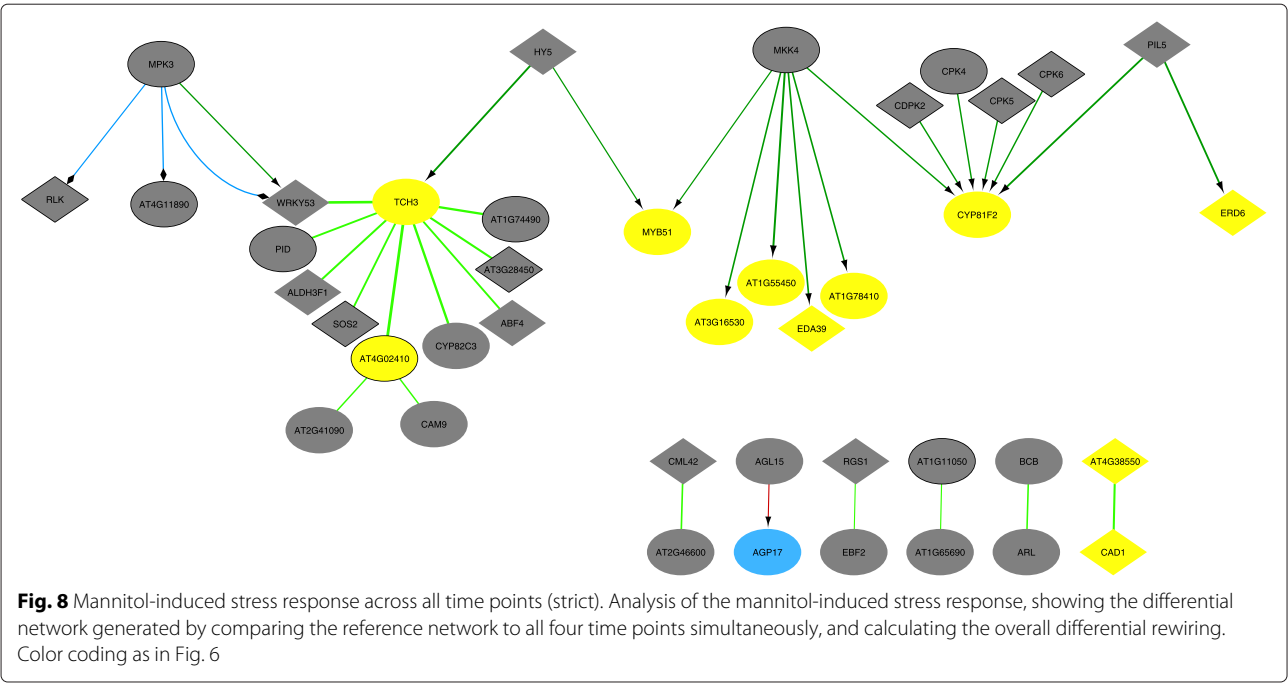
Van Landeghem *et al. BMC Bioinformatics* (2016) 17:18

Page 9 of 12

Fig. 6 Mannitol-induced stress response at 1.5 h. Analysis of the mannitol-induced stress response, depicting the generated differential network at the 1.5 h time point: increase/decrease of regulation in dark green and red respectively, increase/decrease of PPI in light green and orange, increase/decrease in phosphorylation in blue and purple. It is important to note that in these differential networks the arrows point to rewiring events: a decrease of regulation for instance (red arrows) does not necessarily point to an inhibition, but may also indicate a discontinued activation. Diamond nodes represent proteins with a known phosphorylation site, and proteins with a kinase function are shown with a black border. Blue and yellow nodes identify underexpressed and overexpressed genes respectively

Fig. 7 Phenotype of the hy5 mutant on mannitol-stress. Rosettes of WT (*left*) and hy5 mutants (*right*) on control medium (*top* panel) and mannitol-supplemented medium (*bottom* panel). Plants are 22 days old. Scalebar = 1 cm

parallel with other regulators previously identified in the early mannitol-response of growing Arabidopsis leaves [24, 35].

Finally, we can apply a less stringent criterium to the inference of differential networks by only requiring that three out of four time points need to match for a rewiring event to be included in the differential network (Fig. 9). This results in more robust network inference, as the differential network would remain the same when some noise would be introduced at one of the time points. Additionally, this method provides a more complete view on the rewiring pathway occurring in response to osmotic stress in plants. All these different settings and options are also available when generating the differential networks through the Cytoscape plugin.
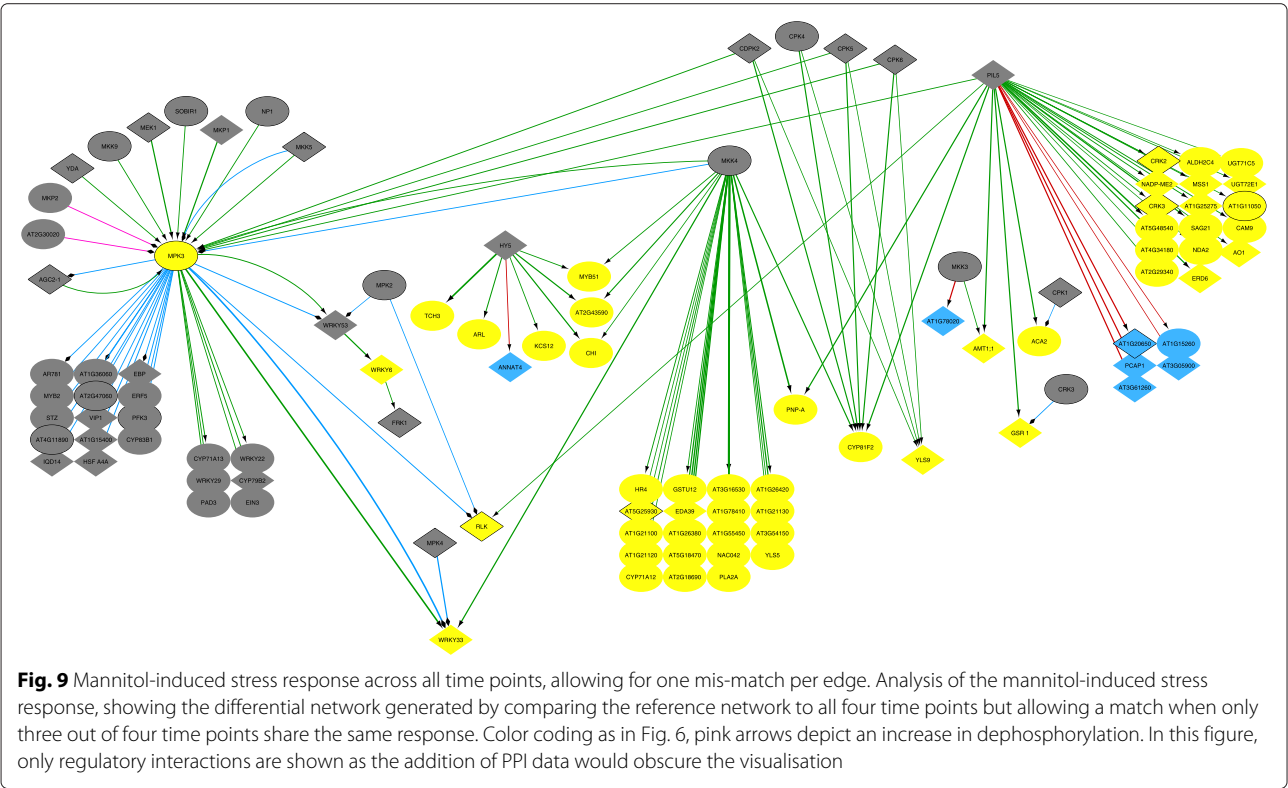
Discussion and conclusion

We have developed an open-source framework, called Diffany, for the inference of differential networks from an arbitrary set of input networks. This input set always contains one reference network which represents the interactome of an untreated/unperturbed organism, while all other networks are condition-specific, each modelling the interactome of the same organism subjected to a specific environmental condition or stimulus. Differential networks allow focusing specifically on the rewiring of

Fig. 8 Mannitol-induced stress response across all time points (strict). Analysis of the mannitol-induced stress response, showing the differential network generated by comparing the reference network to all four time points simultaneously, and calculating the overall differential rewiring. Color coding as in Fig. 6

the network as a response to such stimuli, by modelling only the changed interactions. At the same time, interactions that remain (largely) the same are summarised in a 'consensus' network that provides insight into the basic interactions that are not influenced by changes of internal or external conditions. The analysis of these differential and consensus networks provides a unique opportunity to enhance our understanding of rewiring events occurring for instance when plants undergo environmental stress, or when a disease manifests in the human body.

Fig. 9 Mannitol-induced stress response across all time points, allowing for one mis-match per edge. Analysis of the mannitol-induced stress response, showing the differential network generated by comparing the reference network to all four time points but allowing a match when only three out of four time points share the same response. Color coding as in Fig. 6, pink arrows depict an increase in dephosphorylation. In this figure, only regulatory interactions are shown as the addition of PPI data would obscure the visualisation

Van Landeghem *et al. BMC Bioinformatics* (2016) 17:18

Page 11 of 12

Further, the fact that the framework can compare an arbitrary number of condition-specific networks to one reference network at the same time, forms a powerful tool to analyse distinct but related conditions, such as different human diseases that may share a defected pathway, or various abiotic stresses influencing a plant in a similar fashion.

In comparison to previous work in the emerging field of differential network biology, Diffany is the first generic framework that provides data integration functionality in the context of differential networks. To this end, we have implemented an Interaction Ontology which enables seamless integration of different interaction types, provides semantic interpretation, and deals with heterogeneous input networks containing both directed and symmetrical relations. This ontology forms the backbone for the implementation of the network inference methods that produces differential networks. As in any Systems Biology study or application, a known challenge involves the issue of non-existing edges: an interaction may be missing from the network because it was experimentally determined that no association occurred, or it may simply be that there is a lack of evidence for the interaction, not actually excluding its existence. To deal with these cases, Diffany allows the definition of negated edges, which are explicit recordings of interactions that were determined not to happen under a specific condition.

To provide easy access to the basic functionality of inference and visualisation of differential and consensus networks, we have developed a commandline interface and a Cytoscape plugin. The Cytoscape plugin allows to generate custom differential networks as well as reproduce the use-cases described in this paper. All relevant code is released under an open-source license.

Finally, we have illustrated the practical utility of Diffany on a study involving osmotic stress responses in *Arabidopsis thaliana*. The resulting differential networks were found to be concise and coherent, modelling the response to mannitol-induced stress adequately. The analysis of these differential networks and a preliminary experimental validation has led to the identification of new candidate regulators for early mannitol-response, such as PIL5 and HY5, which likely contribute to the fast transcriptional induction of mannitol-responsive genes. Further detailed biological validation, including for instance ChIP experiments and experimental systems biology approaches, are necessary to confirm the role of HY5 in this context and fully unravel the early stress-induced rewiring events of this complex regulatory network.

Endnote

^aAPI at http://bioinformatics.psb.ugent.be/ supplementary_data/solan/diffany/.

Additional files

> **Additional file 1: Overview of the Diffany framework.** Overview of the Diffany framework and its typical usage in a specific experiment involving the perturbation of an interactome under one or more conditions. (DOCX 183 KB)
>
> **Additional file 2: List of differentially expressed genes.** Dataset of differentially expressed genes, as originally published by [24]. Here, those genes are listed that are differentially expressed in at least one of the 4 time points and in either the more (FDR < 0.05) or less (FDR < 0.1) stringent dataset. This file also depicts the overlap of genes at the different time points. (XLSX 514 KB)
>
> **Additional file 3: Experimental methodology.** Methodological details of the experiments performed on the putative HY5 regulator. (DOCX 22 KB)
>
> **Additional file 4: Figure showing the experimental validation of the putative HY5 regulator.** Detailed analysis of hy5 mutants and WT lines when exposed to mannitol-induced stress, comparing both leaf area as well as expression levels of putative HY5-target genes such as *TCH3* and *MYB51*. (DOCX 472 KB)

Author details
¹Department of Plant Systems Biology, VIB, Technologiepark 927, 9052 Ghent, Belgium. ²Department of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 927, 9052 Ghent, Belgium. ³Genomics Research Institute, University of Pretoria, Private bag X20, 0028 Pretoria, South Africa.

References

1. Srinivasan BS, Shah NH, Flannick JA, Abeliuk E, Novak AF, Batzoglou S. Current progress in network research: toward reference networks for key model organisms. Briefings in Bioinforma. 2007;8(5):318–32. doi:10.1093/bib/bbm038.
2. Balaji S, Babu MM, Aravind L. Interplay between network structures, regulatory modes and sensing mechanisms of transcription factors in the transcriptional regulatory network of E. coli. J Mole Biol;372(4):1108–22.
3. Fiedler D, Braberg H, Mehta M, Chechik G, Cagney G, Mukherjee P, et al. Functional organization of the S. cerevisiae phosphorylation network. Cell. 2009;136(5):952–63.
4. Friedel S, Usadel B, Von Wirén N, Sreenivasulu N. Reverse engineering: A key component of systems biology to unravel global abiotic stress cross-talk. Front Plant Sci. 2012;3(294):1–16. doi:10.3389/fpls.2012.00294.

Van Landeghem *et al. BMC Bioinformatics* (2016) 17:18

Page 12 of 12

5. Przytycka TM, Singh M, Slonim DK. Toward the dynamic interactome: it's about time. Brief Bioinforma. 2010;11(1):15–29. doi:10.1093/bib/bbp057.
6. Ideker T, Krogan NJ. Differential network biology. Mole Syst Biol. 2012;8(565):1–9. doi:10.1038/msb.2011.99.
7. Gill R, Datta S, Datta S. A statistical framework for differential network analysis from microarray data. BMC Bioinforma. 2010;11(1):95. doi:10.1186/1471-2105-11-95.
8. Tesson B, Breitling R, Jansen R. DiffCoEx: a simple and sensitive method to find differentially coexpressed gene modules. BMC Bioinforma. 2010;11(1):497. doi:10.1186/1471-2105-11-497.
9. Bandyopadhyay S, Mehta M, Kuo D, Sung MK, Chuang R, Jaehnig EJ, et al. Rewiring of genetic networks in response to DNA damage. Science. 2010;330(6009):1385–89. doi:10.1126/science.1195618.
10. Bisson N, James DA, Ivosev G, Tate SA, Bonner R, Taylor L, et al. Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. Nature Biotechnol. 2011;29:653–8. doi:10.1038/nbt.1905.
11. Zhang B, Li H, Riggins RB, Zhan M, Xuan J, Zhang Z, et al. Differential dependency network analysis to identify condition-specific topological changes in biological networks. Bioinforma. 2009;25(4):526–32. doi:10.1093/bioinformatics/btn660.
12. Bean G, Ideker T. Differential analysis of high-throughput quantitative genetic interaction data. Genome Biol. 2012;13(12):123. doi:10.1186/gb-2012-13-12-r123.
13. Amar D, Shamir R. Constructing module maps for integrated analysis of heterogeneous biological networks. Nucleic Acids Res. 2014;42(7): 4208–219. doi:10.1093/nar/gku102.
14. Hudson NJ, Reverter A, Dalrymple BP. A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. PLoS Comput Biol. 2009;5(5):1000382. doi:10.1371/journal.pcbi.1000382.
15. Krouk G, Mirowski P, LeCun Y, Shasha D, Coruzzi G. Predictive network modeling of the high-resolution dynamic plant transcriptome in response to nitrate. Genome Biol. 2010;11(12):123. doi:10.1186/gb-2010-11-12-r123.
16. Guan Y, Gorenshteyn D, Burmeister M, Wong AK, Schimenti JC, Handel MA, et al. Tissue-specific functional networks for prioritizing phenotype and disease genes. PLoS Comput Biol. 2012;8(9):1002694. doi:10.1371/journal.pcbi.1002694.
17. Magger O, Waldman YY, Ruppin E, Sharan R. Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. PLoS Comput Biol. 2012;8(9):1002690. doi:10.1371/journal.pcbi.1002690.
18. Ma C, Xin M, Feldmann KA, Wang X. Machine learning-based differential network analysis: A study of stress-responsive transcriptomes in arabidopsis. The Plant Cell Online. 2014;26(2):520–37. doi:10.1105/tpc.113.121913.
19. Novère NL, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The Systems Biology Graphical Notation. Nat Biotechnol;27:735–41.
20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504. doi:10.1101/gr.1239303.
21. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinforma. 2009;25(8): 1091–93. doi:10.1093/bioinformatics/btp101.
22. Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. Bioinforma. 2005;21(16):3448–449. doi:10.1093/bioinformatics/bti551.
23. Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, et al. GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinforma. 2010;26(22):2927–928. doi:10.1093/bioinformatics/btq562.
24. Skirycz A, Claeys H, De Bodt S, Oikawa A, Shinoda S, Andriankaja M, et al. Pause-and-stop: The effects of osmotic stress on cell proliferation during early leaf development in arabidopsis and a role for ethylene signaling in cell cycle arrest. The Plant Cell Online. 2011;23(5):1876–88. doi:10.1105/tpc.111.084160.
25. Irizarry RA, Hobbs B, Collin F, Beazer?Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostat. 2003;4(2):249–64. doi:10.1093/biostatistics/4.2.249.
26. Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):80. doi:10.1186/gb-2004-5-10-r80.
27. Smyth GK. Limma: linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. New York: Springer. p. 397–420.
28. De Bodt S, Carvajal D, Hollunder J, Van den Cruyce J, Movahedi S, Inzé D. CORNET: A user-friendly tool for data mining and integration. Plant Physiology. 2010;152(3):1167–79.
29. De Bodt S, Hollunder J, Nelissen H, Meulemeester N, Inzé D. CORNET 2.0: integrating plant coexpression, protein-protein interactions, regulatory interactions, gene associations and functional annotations. New Phytologist. 2012;195(3):707–20.
30. Yilmaz A, Mejia-Guerra MK, Kurz K, Liang X, Welch L, Grotewold E. Agris: the arabidopsis gene regulatory information server, an update. Nucleic Acids Res. 2011;39(suppl 1):1118–22. doi:10.1093/nar/gkq1120.
31. Zulawski M, Braginets R, Schulze WX. PhosPhAt goes kinases - searchable protein kinase target information in the plant phosphorylation site database PhosPhAt. Nucleic Acids Res. 2013;41(D1):1176–84. doi:10.1093/nar/gks1081.
32. González Besteiro MA, Ulm R. Phosphorylation and stabilization of arabidopsis MAP Kinase Phosphatase 1 in response to UV-B stress. J Biol Chem. 2013;288(1):480–6. doi:10.1074/jbc.M112.434654.
33. Guo H, Ecker JR. Plant responses to ethylene gas are mediated by SCFEBF1/EBF2-dependent proteolysis of EIN3 transcription factor. Cell;115(6):667–77.
34. Potuschak T, Lechner E, Parmentier Y, Yanagisawa S, Grava S, Koncz C, et al. EIN3-dependent regulation of plant ethylene hormone signaling by two arabidopsis F box proteins: EBF1 and EBF2. Cell. 2003;115(6):679–89.
35. Dubois M, Skirycz A, Claeys H, Maleux K, Dhondt S, De Bodt S, et al. ETHYLENE RESPONSE FACTOR6 acts as a central regulator of leaf growth under water-limiting conditions in arabidopsis. Plant Physiology. 2013;162(1):319–32. doi:10.1104/pp.113.216341.