

Guiding the Interactive Exploration of Metabolic Pathway Interconnections

Ilir Jusufi

Linnaeus University, Sweden

Christian Klukas

IPK Gatersleben, Germany

Andreas Kerren

Linnaeus University, Sweden

Falk Schreiber

Martin Luther University
Halle-Wittenberg & IPK
Gatersleben, Germany

Abstract

Approaches to investigate biological processes have been of strong interest in the last few years and are the focus of several research areas, especially Systems Biology. Biochemical networks as representations of processes are very important for a comprehensive understanding of living beings. Drawings of these networks are often visually overloaded and do not scale. A common solution to deal with this complexity is to divide the complete network, for example, the metabolism, into a large set of single pathways that are hierarchically structured. If those pathways are visualized, this strategy generates additional navigation and exploration problems as the user loses the context within the complete network.

In this article, we present a general solution to this problem of visualizing interconnected pathways and discuss it in context of biochemical networks. Our new visualization approach supports the analyst in obtaining an overview to related pathways if they are working within a particular pathway of interest. By using glyphs, brushing, and topological information of the related pathways, our interactive visualization is able to intuitively guide the exploration and navigation process, and thus the analysis processes too. To deal with real data and current networks, our tool has been implemented as a plugin for the VANTED system.

Keywords network visualization, biochemical networks, navigation, glyphs

1 Introduction

The visualization of large and complex networks is still one of the most challenging tasks in graph drawing and information visualization [1]. Although that considerable work on graph visualization has been done and powerful tools exist, some fundamental problems are still not adequately solved. One example of such a problem is the embedding of additional information which is attached to the network nodes and edges. Another example is the well-known scalability problem that arises in trying to visualize huge networks. Drawings of these networks are often visually overloaded and do not scale. To focus on the latter problem, one traditional solution is a hierarchical structuring of the entire network if possible, that is, the complete network is divided into many parts. This is also a common procedure in biochemical network drawings.

Methods for the investigation of biological processes such as metabolism, the regulation of genes, and the interaction of proteins have been of strong interest in the

last few years. Biochemical networks, which are typical representations of these processes, are crucial for a comprehensive understanding of living beings. One example is metabolism, where nodes of the network represent metabolites and reactions, and edges connecting metabolites with reactions and reactions with metabolites represent the ways metabolites can be changed into each other via the respective reactions. The division of a metabolic network into a large set of single (often hierarchically structured) pathways is currently the state-of-the-art strategy, see also Section 2. If these pathways are visualized, then additional navigation and exploration problems occur, because the user loses the overview of the entire network. Similar problems can be identified in other domains, such as in the visualization of communication networks, social networks, etc.

In this article, we present a general approach for navigating between interconnected subgraphs. We discuss it in the context of biochemical networks, that is, in a set of hierarchically structured pathways (groups and sub-groups of pathways). Our visualization approach

initially supports the researcher to obtain an overview of pathways, which are connected to the currently considered pathway via its nodes and edges. Based on this contextual information, our interactive visualization tool is able to guide the exploration and navigation process by using glyphs, brushing, and topological information of the involved pathways. In this way, it can help the analyst to perform the analysis processes more efficiently. To be able to work with real data, current networks and concrete problems, we implemented our approach as a plugin for the VANTED system. This system was originally developed for the visualization and analysis of experimental data in the context of biochemical networks.

Our article is organized as follows: Background information and related work, concerning the biological as well as visualization aspects are presented in Section 2. The underlying VANTED system is briefly illustrated in Section 3, while details of our novel navigation and exploration approach are given in Section 4. Here, we also provide an overview of limitations and implementation issues. A use-case scenario is discussed in the next section followed by a conclusion and ideas for future improvements.

2 Background and Related Work

Biological networks such as gene regulatory networks and metabolic networks play an important role in the investigation of biological processes. They are crucial for a comprehensive understanding of living beings. However, due to the steady progress of knowledge, which results in increasingly large and complex networks, these networks are increasingly difficult to interpret and visualize. An example is network information managed in the KEGG database [2], which contains hierarchically structured pathways with in total more than 10,000 nodes representing genes, proteins/enzymes, and metabolites.

To deal with such large networks several methods and tools have been developed. Especially for the visualization and analysis of biological networks many tools are available from the web, and some wider used software examples include CellDesigner [3], Cytoscape [4], ONDEX [5], Pathway Projector [6], PathVisio [7], VANTED [8], and VisANT [9]. Simple approaches try to visualize the complete network on the screen and use zooming and panning for navigation (examples are common graph drawing or network analysis tools). As

such large networks are not easily understandable, they are often separated into overlapping pathways. Commonly these pathways are structured hierarchically, for example, in metabolism several pathways representing the synthesis of amino acids are grouped into a super-pathway *amino acid metabolism*. Dividing networks into pathways helps in reducing the overall complexity, but has the disadvantage that a user loses the context. An example is the Web-interface of the KEGG database [2], which allows the navigation from one pathway image by clicking onto a link to the next, connected pathway image. However, this approach means that connections to elements in other pathways are hard to follow, and that many of the network elements are multiplied several times in different pathways. Other approaches, such as KGML-ED [10], try to improve the navigation between pathways by providing an overview (top nodes of the hierarchy) and functions to zoom into such top nodes, or by extending the pathway by connected pathways within the same frame. However, the main problem remains, which is that either too much information (e. g., many or all pathways at once) is shown, and therefore the picture is not readable anymore, or that the connection to other pathways is lost or difficult to follow. MetaViz presented by Bourqui *et al.* [11] follows a similar approach. It simultaneously visualizes the global network topology as well as the metabolic pathways without multiplying network nodes. Pathways are drawn by using different graph drawing algorithms. In addition, a user-defined set of focus pathways can be drawn entirely and independently due to specific domain requirements. The interconnection of pathways is displayed by using an orthogonal graph drawing algorithm. This strategy leads to long orthogonal edges, which are often hard to follow. Moreover, the connection overview is lost in case the user needs to zoom into specific pathways. In contrast, our approach provides an overview of the interconnections between a single focus pathway and connected pathways, instead of providing a general overview with additional edges. To the best of our knowledge, biologists usually have no problem with node redundancies. They are more interested in recognizing “their” pathways during the exploration process.

Node duplicates are not exclusively used in context of biochemistry. Henry *et al.* [12] discuss duplications for the purpose of decluttering graph visualizations in clustered social networks. They illuminate different applications when to duplicate and how. Based on a controlled experiment described in their paper, the authors

also present general guidelines for node duplications in clustered graph representations, such as the interactive highlighting of duplicated nodes and links is desirable. This specific issue is addressed by our approach too.

Current approaches are mostly based on close interdisciplinary work between visualization or graph drawing researchers and domain experts, such as in biochemistry. Saraiya *et al.* [13] discussed requirements of biochemical network visualization collected from interviews with domain experts. The authors observed five crucial needs that are important for biologists working on pathway analysis, but still not completely realized in existing visualization tools. Our approach presented in this article realizes a 2D solution for the need to overview “multiple pathways simultaneously with interconnections between them” [14]. Specifically, “incoming and outgoing visual links could enable users to view how other pathways can potentially affect or be affected by the focus pathway at each node. In a densely populated pathway, it is important to be able to analyze connectivity between components” [13]. A good overview on open problems and challenges in biological network visualization is also provided by Albrecht *et al.* [15, 16].

A related project is the Caleydo framework [17] that extends the KEGG pathways drawings into 2.5D, similar to the work of Kerren [18] or Brandes *et al.* [19], combined with brushing, highlighting, focus&context, and detail on demand. In this way, it supports the interactive exploration and navigation between several interconnected components by using the 3rd dimension. Drawbacks of this approach are the restricted number of connected pathways that can be displayed because of the bucket-like visualization setup and the unavoidable clutter/overlaps in 3D if a node-link metaphor is used.

Recently, Rohrschneider *et al.* [20] presented a grid-based visualization approach for metabolic networks supported by a focus&context view. This view is based on a Table Lens method [21], which provides multiple foci. Together with the grid-layout, the user’s mental map [22] is also preserved. But even by using these rich navigation and interaction possibilities the cognitive overload is still high and the approach follows a classical top-down navigation method. In case the analyst starts from a single node, he/she has to follow the links to other pathways successively. Here in this article, we present a technique which strives to improve this situation. Because of space restrictions, we do not give a comprehensive overview of other related tools and approaches and refer to the papers

[16, 20, 23] instead. These papers provide a comprehensive list of related work—not only focused on pathway interconnections—and name navigation and interaction facilities as essential in network visualization.

As already mentioned in the introduction, our interactive exploration and navigation approach is based on the VANTED system, which is briefly presented in the following section.

3 VANTED

VANTED [8, 24] is a flexible system that aims to support researchers in visualizing and analyzing data from experiments in the context of biological networks or pathways. The drawing algorithms of VANTED are based on the extensible graph library and editor Gravisto [25]. VANTED can be easily extended by plugins, and several plugins have been developed, for example, to compute and visualize fluxes in metabolism or to connect databases to VANTED.

The original way of navigating between interconnected hierarchical pathways in VANTED is shown in Figure 1. On the right hand side, a pathway hierarchy is displayed in which the *oxidative phosphorylation* pathway is selected and shown on the left. This pathway is a child of the *energy metabolism* super-pathway and relatively small. Within the pathway drawing, one node *-fumarate-* is in the focus of the user. By mouse-click it is possible to show all links to other pathway nodes connected to *fumarate*. In this example, two rounded rectangles are connected with this node using dashed gray lines, whereas each rectangle represents another pathway. If the user is interested in following, for example, the path to the *TCA cycle* (part of the *carbohydrate metabolism*), then they have to click on the corresponding rounded rectangle to open a new window with the target pathway displayed. All subsequent steps occur analogously.

It is easy to understand that this navigation strategy should be improved, because of the unavoidable clutter if the pathways are large and many nodes are selected. Furthermore, it is not really possible to get an overview of all the pathways linked to the currently selected node(s) and of the overall distribution of all outgoing links.

In this article, we use the MetaCrop database [26] that summarizes diverse information about metabolic pathways in crop plants and allows the automatic export of information for the creation of detailed metabolic

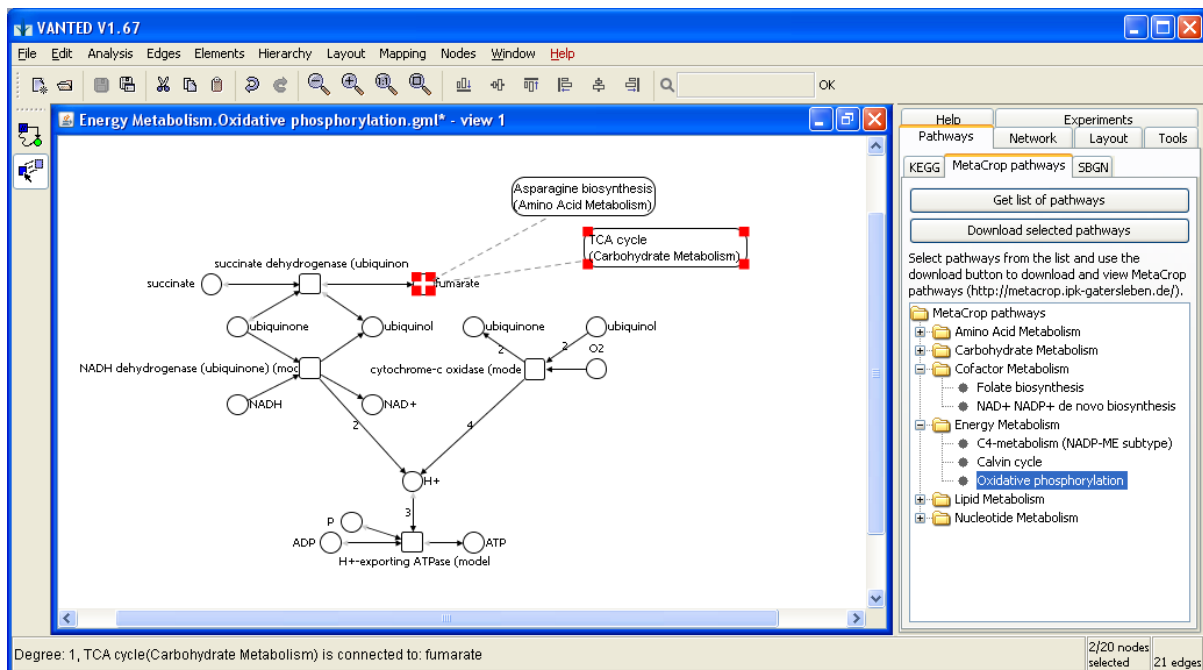


Figure 1: Original navigation approach between pathways provided by VANTED. Shown is a metabolic pathway with circles representing metabolites and rectangles representing reactions. The large rectangles with rounded corners represent links to other pathways. Background coloring to show different cell compartments was switched off to facilitate the identification of the different graphical elements by the reader.

models. MetaCROP is based on the freely available Meta-All system [27] and can be directly accessed by VANTED. In MetaCROP, metabolites are depicted as circular nodes; proteins (enzymes or transporters) are shown as rectangular nodes. In the following screenshot examples, a background coloring is used to visually represent different cell compartments (different parts within the cell, cf. also Figure 4). This information can optionally be used to restrict links to substances represented in multiple pathways to only those cases, where the substance name and the particular compartment information is matching.

4 Novel Approaches to Guide the Exploration Process

Our visualization tool is currently focused on *MetaCROP pathways*, which are placed in six groups (*Amino Acid Metabolism*, *Carbohydrate Metabolism*, *Cofactor Metabolism*, *Energy Metabolism*, *Lipid Metabolism*, and *Nucleotide Metabolism*) consisting of 38 pathways

in total. Thus, the MetaCROP hierarchy has two levels only. The relatively small number of pathways and their more or less distinct graph topology makes these graphs easy to identify by professionals (depending on the graph layout), which is an important aspect for our approach as discussed in this section.

4.1 Navigation Glyphs

In a first step to overcome these problems, we developed small *navigation glyphs*, which are embedded within the network nodes. They show the pathway link situation outgoing from a specific node. Figure 2 shows one instance of a navigation glyph. The motivation behind these rose-shaped diagrams is straightforward as it follows the idea of *Nightingale's rose diagram* [28] (see also the *DataMeadow* [29] approach, for example) and extends it by reflecting the pathway hierarchy. Our approach is similar to the work of Shen *et al.* [30] where attribute values are mapped to the size of the “glyph petals” as well as to the work of Elmquist *et al.* [31] where strength and direction of causal rela-

tionships are visualized by colored polygons. The color coding of the “glyph petals” represents the pathway group (level 1 in the hierarchy), while their size/length shows if there is a link (to the same node in another pathway) or not. In more detail, our sample glyph in Figure 2 is attached to one specific node in a pathway, and this node is connected to five other pathway groups displayed in five different colors. For the color coding of the pathway groups, we use a subset of the standard colors for nominal data types suggested by Ware [32, p. 126]. The highlighted sector gives an impression about the strength of the connectivity to a specific pathway group (*Amino Acid Metabolism* in this case). It is divided into 12 petals in sum, representing all pathways belonging to this group, and 6 longer/extended petals. This means that there are links to 6 sub-pathways of the current group in the second level of the hierarchy. It is possible that such glyphs have gaps in the sense of missing petals as shown in the lower part of the example. This occurs if the underlying VANTED system detects a pathway group that is completely unconnected to the currently considered node.

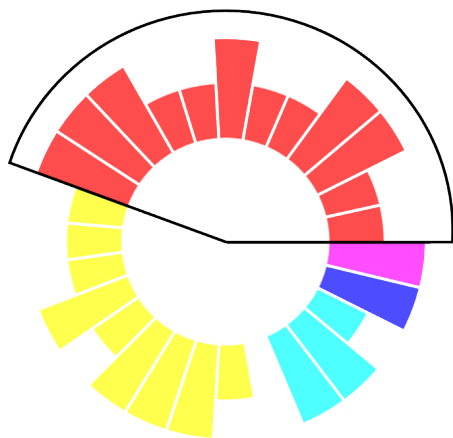


Figure 2: Navigation Glyph. The highlighted sector represents all connected pathways belonging to one super-pathway (or pathway group), e.g., *Amino Acid Metabolism*.

Our navigation glyphs provide a good and simple overview of the relationships to connected pathway groups outgoing from a single node, which is a clear improvement compared to the original approach of VANTED. The navigation is then performed using a right click on a node, which will display a context menu with a list of links to all connected pathways, see Fig-

ure 3. This follows a similar idea as the Bring & Go approach presented by Moscovich *et al.* [33], however without smooth animations or link sliding. Here, a certain pathway window is directly activated after a click on a specific link. In this way, the navigation is comparable to the old approach explained earlier, and it has also similar drawbacks, such as losing a direct link to the previous pathway window: in order to return to the previously shown pathway, we need to minimize the newly opened window and find the first one, or to directly access the node that has a link to it. Moreover, we have no clear insight into all links of the current pathway, because of the missing interactive mapping of the petals to the connected pathways.

4.2 Extended Navigation Approach

Our main idea is to overcome the drawbacks of the navigation glyphs presented in the previous subsection by

1. providing an overview of all connected pathways of the current focus pathway,
2. showing statistical information about the focus pathway’s link distribution, and
3. facilitating the navigation from nodes in the focus pathway to different target pathways.

For this, we introduce new visualization and brushing techniques coupled with seamless interaction of various parts of the system. That will make the navigation of different pathways easier and yields a more intuitive exploration experience. We also investigated the need to provide more insight into the links from a particular pathway by introducing a possibility to compare the number of total references to all connected pathways.

Navigation Window The approach consists of laying out minimized images of pathway drawings, called *graph icons*, around the main view that contains the currently considered focus pathway, see Figure 4. Both together form the *navigation window* of our VANTED plugin. The graph icons are used to navigate to a specific pathway by mouse clicks, which would load the target pathway into the navigation window. Each graph icon has a background color that is brushed accordingly to the color coding of the navigation glyphs within the pathway nodes, thus making it easy to identify the different pathway groups of the MetaCrop data set. Conceptually, our graph icons are similar to the approach

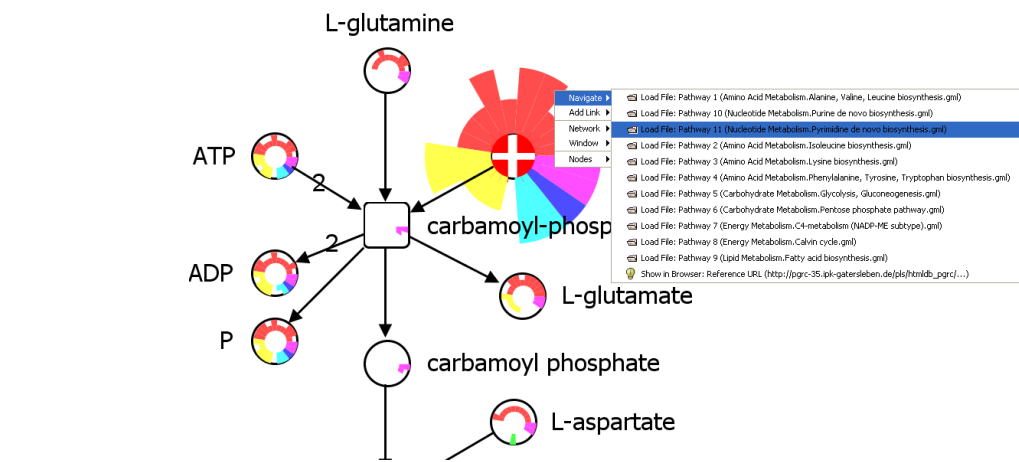


Figure 3: Embedding of our navigation glyphs into some nodes of a (focus) pathway. The glyph of the currently selected node is highlighted and magnified by the system. With a right mouse-click the user is able to open a context menu with a link list from which he/she can jump to connected pathways. The target pathway is then displayed in a new window.

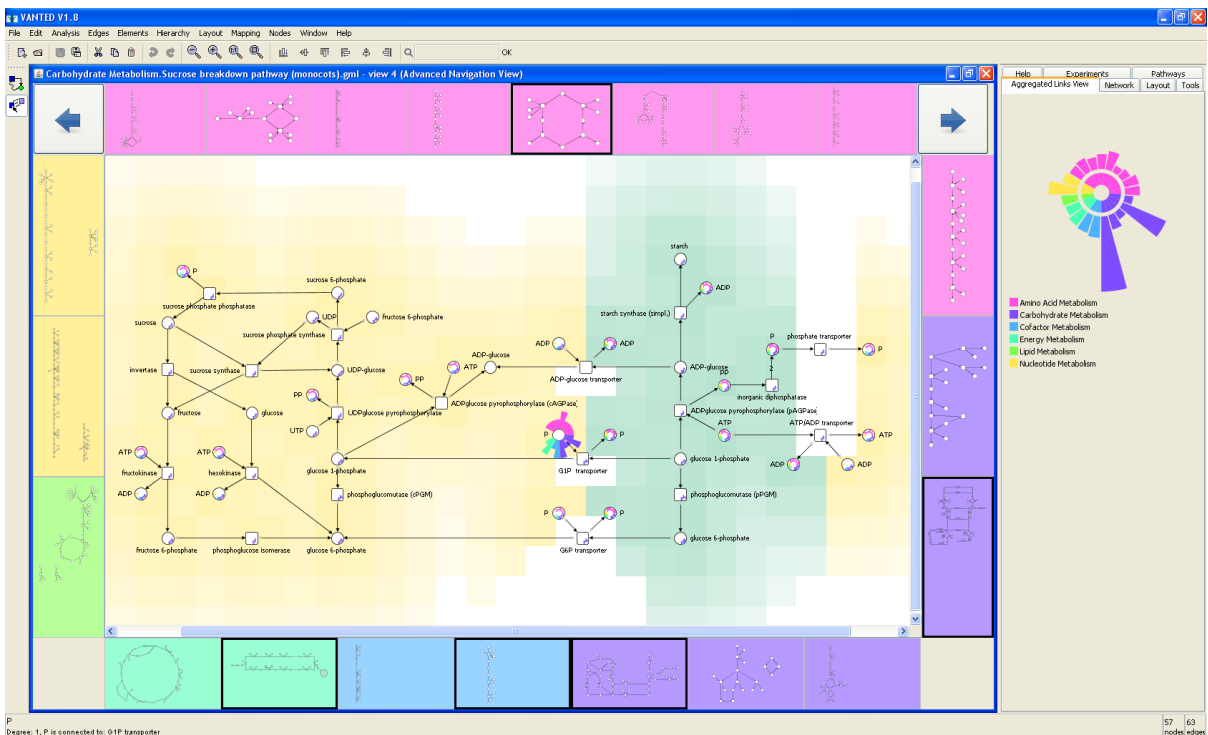


Figure 4: Overview of the new VANTED plugin with the *navigation window* on the left and the *aggregated links view* on the right. The background coloring of the shown focus pathway represents different cell compartments. One node is selected and its navigation glyph was magnified by the system. The glyph petals are directed to the surrounding graph icons. Links to other pathways are double-coded by the longer petals as well as by the highlighted graph icons. The *Link Rose Diagram* in the aggregated links view (on the right) is explained in Section 4.2.

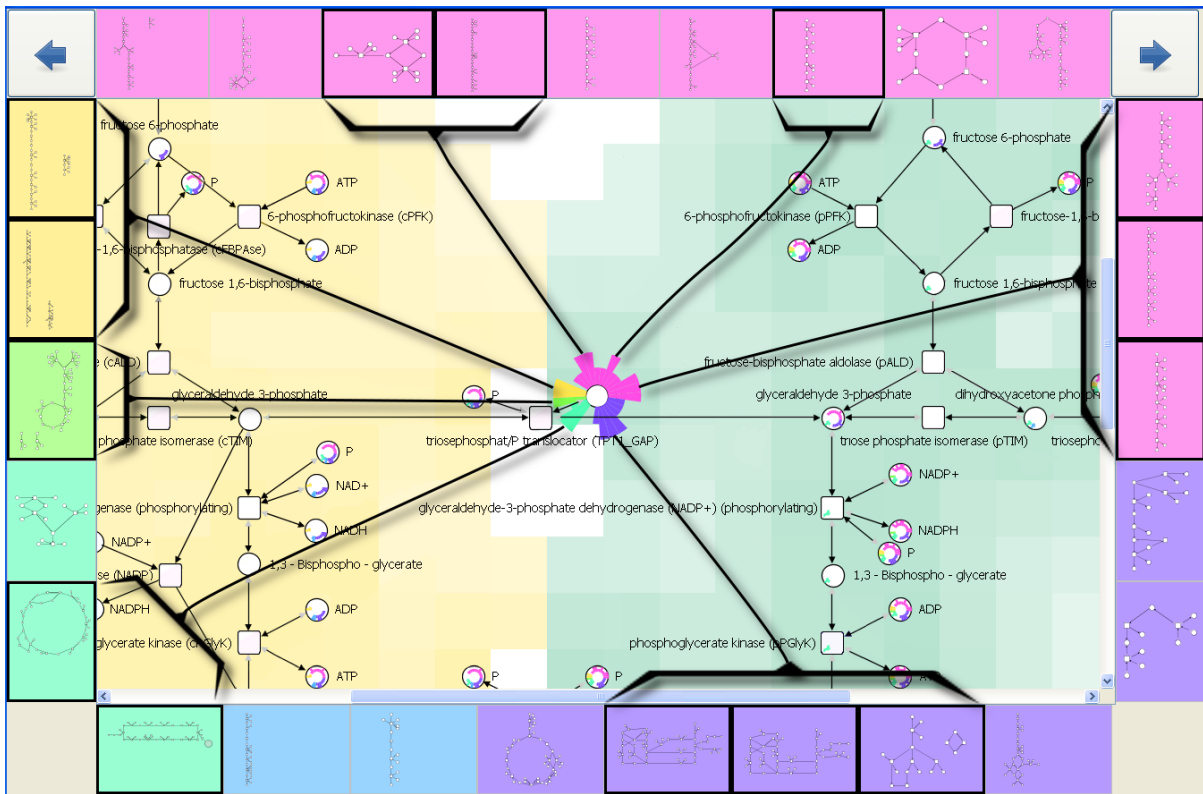


Figure 5: A mouse-over action on a specific node highlights the correlated graph icons (black framed) around the focus pathway. Note that the half-transparent shapes are hand made and are not part of our plugin; they are just drawn to demonstrate the ordering and layout of the approach.

presented by Auber *et al.* [34] or to the subtree icons in the SpaceTree tool [35]. The user is able to recognize the pathways connected to the nodes of the focus pathway based on their graph topology, additionally supported by the aforementioned color coding. As previously stated, this recognition of the target pathways is of particular importance as the user can directly identify well-known pathways with a recognizable and unique layout, such as the *TCA cycle*. The pathway layout is either given (e.g., by information in the database) or can be computed with special algorithms which take additional pathway information into account [36, 37, 38]. For precise information about the pathways, the user can use tooltips showing the name of the pathway and of the pathway group. They appear based on a mouse-over event of a specific graph icon.

The position of the graph icons roughly corresponds to the position of the petals within the small *navigation glyphs*, which in turn provides additional insight to

this existing feature. We chose this layout design since our users already used the navigation glyph metaphor, that is, our solution complements the old one aimed at specific pathway nodes as discussed in Subsection 4.1, while making it more intuitive to use. We consider this to be a better alternative compared to solutions such as making a simple link list that in turn might require scrolling efforts to fit all the links and does not provide a mapping of the pathways with our navigation glyphs. As we will explain later in this article, we have also implemented an additional visualization view (the aggregated links view) that uses the same layout and navigation concept and provides an integrated tool for link aggregation to facilitate the guided interactive exploration.

Using the graph icons, it is possible to navigate through all related pathways in one window, while providing a good overview of all pathway links. Users can easily access all pathways linked from an interesting node through a mouse-over, which will highlight the

corresponding graph icon, cf. Figure 5. This feature allows users to quickly navigate and/or to gain insight into the connectivity of desired pathway components. Additionally, the user might use *back and forward buttons* to facilitate the navigation. They can be found on the left and right upper corner of the navigation window. These buttons will also highlight a specific graph icon based on a mouse-over action, giving an additional insight into the destination at which they would arrive in case of clicking the Back- or the Forward-button, respectively. This helps to keep track of navigation processes and avoids a search for the previous pathway in multiple open windows.

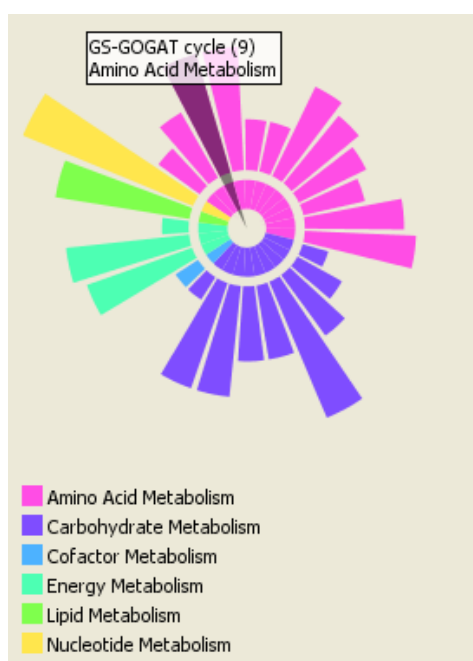


Figure 6: *Link rose diagram*. The “GS-GOGAT cycle” is selected triggering a specific petal to be shadowed in a more saturated color. A mouse-over action shows the name of the pathway in the first line of a tooltip, while the number in brackets reveals the total number of links from the focus pathway to the selected pathway. The second line of the tooltip shows the name of the MetaCrop pathway group that the selected pathway belongs to. Users can instantly notice strongly and/or loosely connected pathways. The legend below helps in identification of different pathway groups. The color-coding of the groups is persistent for specific group, no matter where we navigate to.

Aggregated Links View The graph icons provide an overview of all pathways linked to the current focus pathway, but they cannot show the frequency a particular target pathway is being referenced from the nodes of the focus pathway. Accordingly, a statistical data visualization is desired. Instead of integrating it into the main window, we decided to use a *multiple views* approach that is a widely used technique to visualize different kinds of data simultaneously. This method provides the possibility of choosing the most suitable visualization technique for a specific view and data set [39]. Guided by these ideas we added a separate interactive diagram placed at the upper right corner in a tab, called *aggregated links view*, to present this statistical data. Figure 6 shows a screenshot excerpt of this diagram. As it resembles the navigation glyphs used in the focus pathway and is able to guide the user similarly to a traditional compass rose metaphor, we call it a *link rose*. In contrast to the navigation glyphs that represent the links of a particular pathway node, it displays statistical data about *all* links to other connected pathways in context of the current focus pathway. Thus, it enables users to compare all outgoing pathway links and to show which pathways are strongly connected to the focus pathway.

Mouse-over actions on a rose petal show a tooltip with the name of the corresponding pathway, followed by the number of nodes linked to that specific pathway. The tooltip also presents the pathway group name in the second text line. This gives an additional, textual dimension of insight into the connectivity of a specific graph to other pathways. Following the idea of several coordinated views [40], a click on a rose petal has also an effect on other parts of our visualization approach (besides highlighting the corresponding petal), namely that all graph nodes referencing the selected pathway are also highlighted in the navigation window. This possibility enables users to immediately identify components connected to a particular pathway, and thus, enhances the comprehension of interconnections of different graph components and provides a better overview of the entire network structure. Moreover, if we subsequently navigate to any target pathway, all shared nodes (clones) will be selected in the target pathway consequently revealing the common components between these pathways. A more detailed navigation and exploration example is given in Section 5.

4.3 Scalability

Our visualization and interaction approach scales well when using the MetaCrop database consisted of 38 pathways. We are aware that the graph icons discussed in Section 4.2 depend on size and resolution of the screen as well as on the number of pathway links being visualized. Theoretically, if the display size and resolution decreases or the number of pathways is large enough, our graph icons would be minimized so much that it would be hard to recognize them. Surely, there are various technical solutions to overcome this problem. For example, the system could provide more filtering and/or prioritization features in order to reduce the number of displayed graph icons. The user might not be interested in a specific pathway group, such as *Amino Acid Metabolism*, and deselect those. Then, our tool could delete all target graph icons belonging to the deselected pathway group from the current view. Another way to address the problem of many pathway links is to spatially arrange a fixed rectangular block for each pathway group around the target pathway. Within this block, we could integrate graph icons of different size following a semantic fisheye metaphor. The size could be proportional to the number of matching nodes between the individual target pathways and the focus pathway. However under normal circumstances, we do not need such techniques currently, as described in the following.

Larger data sets, such as KEGG [2] which contains the largest collection of pathways and also both signaling and metabolic pathways, can demonstrate the scalability of our approach. Therefore, we analyzed KEGG focusing on substances (the most commonly shared elements between different pathways) and found that most of the pathways (about 95%) have less than 62 links to other pathways in practical analysis sessions, as shown in the Figure 7. With these findings in mind, we can conclude that the application scales without any difficulties for KEGG in most cases: we used a laptop with a screen resolution of 1280x800 pixels and loaded a pathway with 81 links, which is the extreme case in Figure 7. We found that the graph icons scale even in such extreme cases, although the individual nodes and edges in some of the graph icons were less distinguishable. Keeping in mind that nowadays standard office displays have even larger resolutions, we do not expect any scalability problems for applications in biology. This also holds for the order and size of graphs, i.e., the number of nodes and edges: a typically analyzed (focus) path-

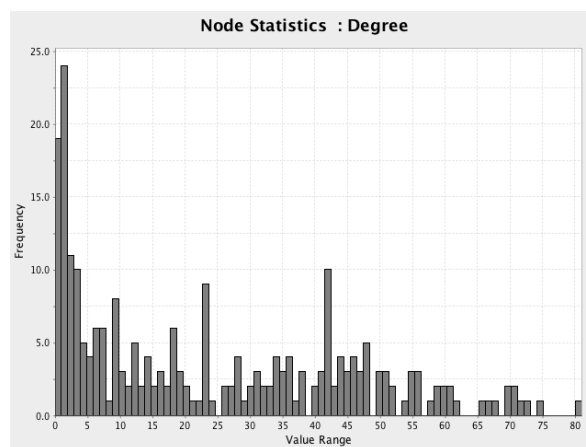


Figure 7: The figure shows the number of links to other pathways (same substance in other pathways) for the reference pathways of KEGG (“Frequency” means number of pathways, e.g., there are 10 pathways which have connections to 42 pathways each). The largest number of links has the pathway *Inositol phosphate metabolism*, with substances of this pathway found in 81 other pathways.

way contains at most a few hundred nodes and edges, which can be easily drawn and interactively displayed by the VANTED system.

As mentioned before, we have also tested our approach with the KEGG data set. Currently, the mapping of KEGG hierarchies/groups is not completely implemented. There are two potential limitations: deeper hierarchies (i.e., more levels) and/or broader hierarchies (i.e., many different groups). The fact that each pathway group has a fixed position around the navigation window facilitates the perception additionally. This idea does not help with deeper levels. However, there are ways to overcome these constraints and we present them in the future work section (Section 6).

4.4 Implementation and Performance

The VANTED system offers a comfortable way to implement various *plugins* as separate projects, see Figure 8 for an overview of its architecture. Plugins can also be easily added by users later and allow a great deal of flexibility in the implementation process and possibilities for building various improvements as well as the embedding of new (visualization) techniques or tools. By utilizing this approach, the users of the VANTED

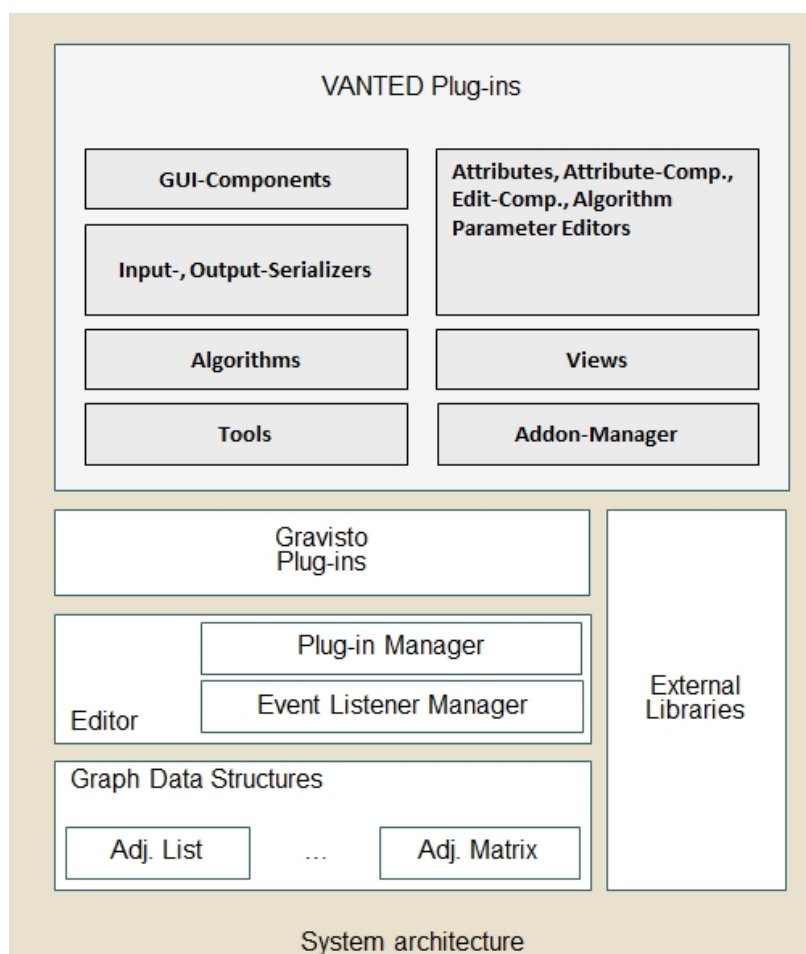


Figure 8: The architecture of the VANTED system. The lower part contains the base functionality provided by Gravisto, the top part represents VANTED plugins for different functions, such as import/export, algorithms and views.

system are able to customize the functionality according to their needs. Our current visualization tool is separated from the original VANTED core implementation as an individual package, i.e., a JAR file that can be imported into the main application as a plugin. VANTED provides an own GUI in order to facilitate the integration of different plugins. Thus, users could only download plugins and import them through a specific form in the VANTED system.

However, some interventions on VANTED's core libraries were necessary in order to implement all our desired features. Particularly the changes in the main view of the navigation window (i.e., where the drawing of the focus pathway is located) required more knowledge of the core VANTED source code. The graph

layout in the main view together with its background coloring and the navigation glyphs (Figure 2) is handled by VANTED itself. In order to lay out the small graph icons around the main view, we had to add several frames. Here, we used simple buttons for representing the graph icons. VANTED provides methods, such as scaling or changing the background color, which allow the small graph icons to be embedded into the buttons. This simplified the implementation considerably as we did not need to compute a layout of the embedded (target) pathways. Additionally, this strategy helped to keep a consistent visual appearance of the specific graphs, whether they are focused in the main view or presented as a graph icon around the main view.

VANTED also supports a relatively easy access to different methods and data structures used to store and retrieve the data as well as their relations to each other. Outgoing from this point, the implementation of the navigation features was simply done via action events of already existing VANTED methods. Using similar resources, brushing and different color codings were implemented. The link rose diagram (Figure 6) was created in a separate view, but it is seamlessly integrated with other parts of the system, again mostly using existing VANTED resources. The visualizations (navigation glyphs (standard and enlarged) and the link rose diagram) were implemented straightforward and are not based on any specific algorithms.

We have tested our VANTED plugin using a Dell Latitude E6400 laptop with a 2.53 GHz Intel Core 2 Duo processor and 3GB of RAM running Windows XP. The time to download the data set depends on the Internet connection and the size/number of the desired pathways. Once all pathways are completely downloaded, users are able to load a focus pathway into the navigation window in 1-2 seconds in worst case. This includes the computation of the link connections to other pathways and the generation of the aggregated links view. Additionally, it might take at most 2-3 seconds to generate the graph icons in case the focus pathway is strongly connected (40-50 links) to target pathways. Based on Figure 7, we can safely conclude that most interactions will occur immediately. Worst case situations occur within the *unit task* level [41].

5 Use-Case Scenario

In this section we will present the way our plugin is used. Usually, users download the pathways they are interested in or the whole MetaCrop data set. In standard visualization approaches the navigation and exploration gets more complicated if the number of pathways being explored increases, and simultaneously the contextual information about the interconnectivity of different components is lost due to the reasons explained in previous sections of the article.

After activating our plugin, users will be able to gain insight into the interconnectivity immediately. The graph icons arranged around the focus pathway are not merely links to other pathways as they provide an overview of the overall connectivity of the focus pathway (cf. Figure 4). Our aggregated links view provides more specific information about the frequency of the

connections to all target pathways. These views together with various interaction and brushing techniques will guide the users to interactively explore the pathways. If a certain petal of the link rose is relatively long compared to the others, we might conclude that the corresponding pathway shares a considerable amount of components with our focus pathway currently displayed in the navigation window. We will explain the navigation process by means of Figure 9. Let us assume that our user has navigated to the *Pentose phosphate pathway* and noticed (cp. the upper part of the figure) that the link rose has a petal that is longer than the others. After a mouse-over event, they discover that the name of the pathway is *Calvin cycle* and that there are 16 links to that particular pathway. If they click on the rose petal, all the nodes referencing that pathway will be selected. This strong interconnection between *Carbohydrate Metabolism* (to which *Pentose phosphate pathway* belongs) and *Energy Metabolism* (to which *Calvin cycle* belongs) can be further investigated by selecting the *Calvin cycle*.

After clicking the graph icon, the user navigates to the desired target pathway (see the bottom part of Figure 9). Previously selected nodes will remain selected if they exist in the new navigated pathway. We notice that several nodes in the *Calvin cycle* are now selected. This example demonstrates that the plugin makes it rather easy to identify the connectivity context of various pathways and facilitates the navigation between them.

Another investigation could start with the *Sucrose breakdown pathway (monocots)*. Again, the link rose has a petal that is much longer than the others. After a mouse-over event, the user discovers that the name of the pathway is *Sucrose breakdown pathway (dicots)* and that there are 54 links to that particular pathway. Clicking on the rose petal selects all the nodes referencing that pathway. In this case, only three nodes are not selected, which clearly shows that this pathway is strongly interconnected with the selected one. After clicking the graph icon and navigating to the desired target pathway *Sucrose breakdown pathway (dicots)* a biologist will immediately recognize that we are navigating in our sample through the same metabolic pathway. In this case it shows the differences between an important pathway in monocotyledons (monocots) and dicotyledons (dicots).

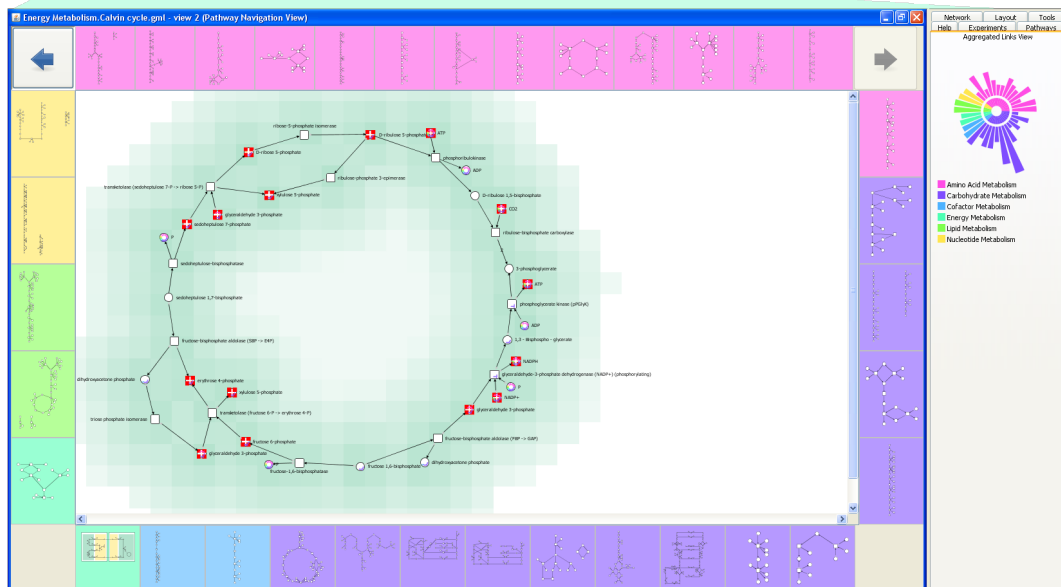
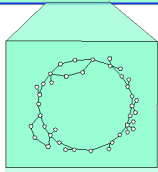
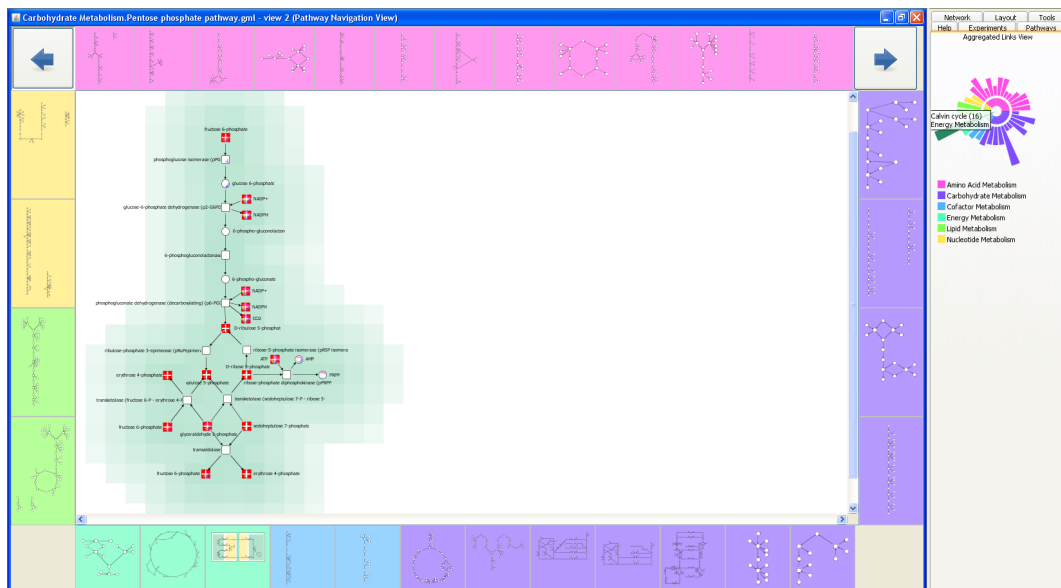


Figure 9: Navigation between two pathways. The selection of the nodes is persistent even when navigating in the new pathway. A feature that helps to discover shared components.

5.1 Feedback from Biologists

We discussed our plugin in three user sessions with four domain experts. These experts are biologists and biochemists used to working with metabolic pathways and familiar with pathway databases such as KEGG and MetaCrop, as well as with the VANTED system. After a short introduction in the usage of the plugin and the idea of the navigation method (introduction was less than five minutes), we asked the experts to work with the plugin to investigate the metabolic pathways stored in the MetaCrop system and report positive and negative aspects.

In general the feedback was very positive. Users described the navigation through pathways as intuitive and very helpful, even as “cool”. The used colors and the size of the glyphs were considered appropriate, the speed of navigation (e. g., replacement of one pathway by the next one) mostly as fast enough. Especially the interaction between the *link rose* and the pathway was considered useful for selecting nodes occurring in another pathway.

However, there were also several suggestions by the domain experts. Although most liked the concept of presenting only those pathways as graph icons which are connected to the current pathway of interest, one user would prefer to have all other pathways (also currently not connected ones) as graph icons placed around the current pathway. Users familiar with the MetaCrop pathways had no problem in navigating to specific pathways (graph icons), but some recommended nevertheless including the name of the pathway in the icon (currently the name is only shown after moving the mouse over the icon). Finally, one user recommended optionally restricting linkage of substances and avoiding co-substances (e. g., ATP, H₂O). This would considerably reduce the number of interconnected pathways.

The users found the method very helpful in navigating through pathways but recommended additional functionality to make it more useful for their work such as the ability to map measured data onto nodes or to see changes in fluxes over reactions. It should be noted that these are aspects which are supported by the VANTED system (e. g., fluxes can be represented by different edge width and additional data such as metabolomics data by additional coloring of nodes or including of diagrams into nodes) and are already used by some of the experts for their data analysis. Such functionality is already possible with our plugin, however, as it is not fo-

cus of this study, we did not extend the discussion with the users into this direction.

6 Conclusion and Future Work

We presented in this article a solution to guide the navigation and exploration process of interconnected pathway visualization and discussed it in the context of biochemical networks. Our visualization approach supports the analyst in obtaining an overview of connected pathways if they are working within the currently focused pathway. This problem appears simple, but is still not sufficiently solved in existing visualization tools.

Our main contribution is to extend a complex system used in practice with novel and intuitive visualization and interaction techniques to guide the domain expert during the exploration and navigation process. Our tool uses well-known techniques, such as glyphs [42, 43], brushing [44], and topological information of the target pathways (realized by the graph icons), to provide the analyst with more meaningful contextual information in hierarchically structured biochemical pathways. We published our plugin, called GLIEP (Glyph-based LInk Exploration of Pathways), on SourceForge [45], where users can easily download it, and programmers can use our source code for further improvements or for adding additional features.

As future work, we plan to improve different issues regarding the visualization and interaction possibilities as well as the availability of our VANTED plugin. They are listed in the following:

- Better solution for the transition between focus and selected context pathway to preserve the user’s mental map according to the work of Misue *et al.* [22].
- Recent research results in large graph exploration, such as the DOI-approach of van Ham and Perer [46], could be interesting for our system too. We plan to analyze the possibility of integrating such or similar techniques.
- A limitation mentioned before regards the visualization of data sets with a deeper level of hierarchies. Currently, we can only support two-level hierarchies. However, this is enough for the discussed data sets. Theoretically, our approach could be extended by using the sunburst approach pre-

sented by Stasko and Zhang [47] or Richard *et al.* [48].

- The link rose diagram could be extended by additional statistical data. The color-coding of the rose diagram and the navigation glyphs could be improved as well.
- Our small usability study with biologists helped us to identify some potential improvements. One suggestion aimed to show all pathways as graph icons including pathways that are currently not connected. We will analyze this proposal, but it would probably lead to a overloaded user interface. Another improvement is to embed the name of a pathway into its graph icon as well as to provide some filtering techniques to reduce the number of interconnected pathways.
- VANTED is a powerful tool used by many domain experts. Thus, we are convinced that our plugin can facilitate their work. Nevertheless, a carefully designed larger user study or evaluation might detect more concrete or conceptual problems of our approach, which can be improved in a successive stage. A comparative study with MetaVis [11] is desirable as well.

References

- [1] Robert S. Laramée and Robert Kosara. Challenges and Unsolved Problems. In Kerren *et al.* [49], pages 231–254.
- [2] M. Kanehisa, S. Goto, S. Kawashima, and A. Nakaya. The KEGG databases at GenomeNet. *Nucleic Acids Research*, 30(1):42–46, 2002.
- [3] A. Funahashi, M. Morohashi, and H. Kitano. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico*, 1(5):159–162, 2003.
- [4] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003.
- [5] J. Köhler, J. Baumbach, J. Taubert, M. Specht, A. Skusa, A. Rüegg, C. Rawlings, P. Verrier, and S. Philippi. Graph-based analysis and visualization of experimental results with ONDEX. *Bioinformatics*, 22(11):1383–1390, 2006.
- [6] N. Kono, K. Arakawa, R. Ogawa, N. Kido, K. Oshita, K. Ikegami, S. Tamaki, and M. Tomita. Pathway projector: Web-based zoomable pathway browser using KEGG atlas and Google maps API. *PLOS One*, 4(11):e7710, 2009.
- [7] M. van Iersel, T. Kelder, A. Pico, K. Hanspers, S. Coort, B. Conklin, and C. Evelo. Presenting and exploring biological pathways with PathVisio. *BMC Bioinformatics*, 9:399, 2008.
- [8] B. H. Junker, C. Klukas, and F. Schreiber. VANTED: A system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics*, 7:109, 2006.
- [9] Z. Hu, J. Mellor, J. Wu, T. Yamada, D. Holloway, and C. DeLisi. VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Research*, 33:W352–W357, 2005.
- [10] C. Klukas and F. Schreiber. Dynamic exploration and editing of KEGG pathway diagrams. *Bioinformatics*, 23(3):344–350, 2007.
- [11] Romain Bourqui, Vincent Lacroix, Ludovic Cottret, David Auber, Patrick Mary, Marie-France Sagot, and Fabien Jourdan. Metabolic network visualization eliminating node redundancy and preserving metabolic pathways. *BMC Systems Biology*, 1:29, 2007.
- [12] Nathalie Henry, Anastasia Bezerianos, and Jean-Daniel Fekete. Improving the readability of clustered social networks using node duplication. *IEEE Transactions on Visualization and Computer Graphics*, 14:1317–1324, 2008.
- [13] Purvi Saraiya, Chris North, and Karen Duca. Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. *Information Visualization*, 4(3):191–205, 2005.
- [14] Mircea Lungu and Kai Xu. Biomedical Information Visualization. In Kerren *et al.* [49], pages 311–342.
- [15] Mario Albrecht, Andreas Kerren, Karsten Klein, Oliver Kohlbacher, Petra Mutzel, Wolfgang Paul,

- Falk Schreiber, and Michael Wybrow. A graph-drawing perspective to some open problems in molecular biology. Technical report, Lehrstuhl XI für Algorithm Engineering, Fakultät für Informatik, Technische Universität Dortmund, 2008.
- [16] Mario Albrecht, Andreas Kerren, Karsten Klein, Oliver Kohlbacher, Petra Mutzel, Wolfgang Paul, Falk Schreiber, and Michael Wybrow. On open problems in biological network visualization. In *Proc. International Symposium on Graph Drawing (GD '09)*, volume 5849 of *LNCS*, pages 256–267. Springer, 2010.
- [17] Alexander Lex, Marc Streit, Ernst Kruijff, and Dieter Schmalstieg. Caleydo: Design and evaluation of a visual analysis framework for gene expression data in its biological context. In *2010 IEEE Pacific Visualization Symposium (PacificVis)*, pages 57–64, Taipei, Taiwan, 2010.
- [18] Andreas Kerren. Interactive visualization and automatic analysis of metabolic networks – a project idea. Technical report, Institute of Computer Graphics and Algorithms, Vienna University of Technology, Austria, 2003.
- [19] U. Brandes, T. Dwyer, and F. Schreiber. Visual understanding of metabolic pathways across organisms using layout in two and a half dimensions. *Journal of Integrative Bioinformatics*, 1(1):119–132, 2004.
- [20] Markus Rohrschneider, Christian Heine, André Reichenbach, Andreas Kerren, and Gerik Scheuermann. A novel grid-based visualization approach for metabolic networks with advanced focus&context view. In *Proc. International Symposium on Graph Drawing (GD '09)*, volume 5849 of *LNCS*, pages 268–279. Springer, 2010.
- [21] Ramana Rao and Stuart K. Card. The table lens: merging graphical and symbolic representations in an interactive focus+context visualization for tabular information. In *CHI '94: Conference companion on Human factors in computing systems*, page 222. ACM, 1994.
- [22] K. Misue, P. Eades, W. Lai, and K. Sugiyama. Layout adjustment and the mental map. *Journal of Visual Languages and Computing*, 6:183–210, 1995.
- [23] Ivan Herman, Guy Melançon, and M. Scott Marshall. Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 6:24–43, 2000.
- [24] C. Klukas and F. Schreiber. Integration of -omics data and networks for biomedical research. *Journal of Integrative Bioinformatics*, 7:112, 2010.
- [25] C. Bachmaier, F. J. Brandenburg, M. Forster, P. Holleis, and M. Raitner. Gravisto: Graph visualization toolkit. In J. Pach, editor, *Proc. International Symposium on Graph Drawing (GD'04)*, volume 3383 of *LNCS*, pages 502–503. Springer, 2005.
- [26] Eva Grafahrend-Belau, Stephan Weise, Dirk Koschützki, Uwe Scholz, Björn H. Junker, and Falk Schreiber. MetaCrop - a detailed database of crop plant metabolism. *Nucleic Acids Research*, 36:D954–D958, 2008.
- [27] S. Weise, I. Grosse, C. Klukas, D. Koschützki, U. Scholz, F. Schreiber, and B. H. Junker. Meta-All: a system for managing metabolic pathway information. *BMC Bioinformatics*, 7:465, 2006.
- [28] F. Nightingale. *Notes on Matters Affecting the Health, Efficiency, and Hospital Administration of the British Army*. Harrison & Sons., 1858.
- [29] Niklas Elmqvist, John Stasko, and Philippas Tsigas. DataMeadow: A Visual Canvas for Analysis of Large-Scale Multivariate Data. *2007 IEEE Symposium on Visual Analytics Science and Technology*, pages 187–194, 2007.
- [30] Zeqian Shen and Kwan-Liu Ma. Mobivis: A visualization system for exploring mobile data. In *Proceedings of IEEE Pacific Visualization Symposium*, pages 175–182. IEEE VGTC, 2008.
- [31] Niklas Elmqvist and Philippas Tsigas. Causality visualization using animated growing polygons. *IEEE Symposium on Information Visualization 2003*, 2003:189–196, 2003.
- [32] Colin Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann, 2nd edition, 2004.

- [33] Tomer Moscovich, Fanny Chevalier, Nathalie Henry, Emmanuel Pietriga, and Jean-Daniel Fekete. Topology-aware navigation in large networks. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 2319–2328, New York, USA, 2009. ACM.
- [34] David Auber, Yves Chiricota, Fabien Jourdan, and Guy Melançon. Multiscale visualization of small world networks. In *IEEE Symposium on Information Visualization (InfoVis '03)*, pages 75–81, Los Alamitos, CA, USA, 2003. IEEE Computer Society.
- [35] C. Plaisant, J. Grosjean, and B.B. Bederson. Spacetree: supporting exploration in large node link tree, design evolution and empirical evaluation. In *IEEE Symposium on Information Visualization (InfoVis '02)*, pages 57–64, Los Alamitos, CA, USA, 2002. IEEE Computer Society.
- [36] P. D. Karp and S. M. Paley. Automated drawing of metabolic pathways. In H. Lim, C. Cantor, and R. Bobbins, editors, *Proc. International Conference on Bioinformatics and Genome Research*, pages 225–238, 1994.
- [37] F. Schreiber. High quality visualization of biochemical pathways in BioPath. *In Silico Biology*, 2(2):59–73, 2002.
- [38] F. Schreiber, T. Dwyer, K. Marriott, and M. Wybrow. A generic algorithm for layout of biological networks. *BMC Bioinformatics*, 10:375.1–12, 2009.
- [39] Carsten Görg, Mathias Pohl, Ermir Qeli, and Kai Xu. Visual Representations. In Kerren et al. [49], pages 163–230.
- [40] Jonathan C. Roberts. Exploratory visualization with multiple linked views. In Alan MacEachren, Menno-Jan Kraak, and Jason Dykes, editors, *Exploring Geovisualization*, chapter 8, pages 159–180. Elseviers, 2004.
- [41] Stuart Card, Jock Mackinlay, and Ben Shneiderman, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, 1999.
- [42] Matthew O. Ward. Xmdvtool: integrating multiple methods for visualizing multivariate data. In *VIS '94: Proc. Conference on Visualization '94*, pages 326–333, Los Alamitos, USA, 1994. IEEE Computer Society Press.
- [43] Matthew O. Ward. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1:194–210, 2002.
- [44] Robert Spence. *Information Visualization: Design for Interaction*. Prentice Hall, 2nd edition, 2007.
- [45] GLIEP. Project homepage, 2010.
- [46] Frank van Ham and Adam Perer. Search, show context, expand on demand: Supporting large graph exploration with degree-of-interest. *IEEE Transactions on Visualization and Computer Graphics*, 15:953–960, 2009.
- [47] John Stasko and Eugene Zhang. Focus+context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *INFOVIS '00: Proc. IEEE Symposium on Information Visualization 2000*, page 57, Washington, USA, 2000. IEEE Computer Society.
- [48] John Stasko, Richard Catrambone, Mark Guzdial, and Kevin McDonald. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies*, 53:663–694, 2000.
- [49] Andreas Kerren, Achim Ebert, and Jörg Meyer, editors. *Human-Centered Visualization Environments*. LNCS Tutorial 4417. Springer, 2007.