

Information Visualization

<http://ivi.sagepub.com/>

Semi-automatic drawing of metabolic networks

Peter Droste, Wolfgang Wiechert and Katharina Nöh

Information Visualization 2012 11: 171 originally published online 22 August 2011

DOI: 10.1177/1473871611413565

The online version of this article can be found at:

<http://ivi.sagepub.com/content/11/3/171>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Information Visualization* can be found at:

Email Alerts: <http://ivi.sagepub.com/cgi/alerts>

Subscriptions: <http://ivi.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://ivi.sagepub.com/content/11/3/171.refs.html>

>> [Version of Record](#) - Jun 27, 2012

[OnlineFirst Version of Record](#) - Aug 22, 2011

[What is This?](#)

Semi-automatic drawing of metabolic networks

Peter Droste¹, Wolfgang Wiechert¹ and Katharina Nöh¹

Information Visualization
11(3) 171–187
© The Author(s) 2011
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1473871611413565
ivi.sagepub.com



Abstract

In the living cell, biochemical reactions catalyzed by enzymes are the drivers for metabolic processes like growth, energy production, and replication. Metabolic networks are the representation of these processes describing the complex interactions of biochemical compounds. The large amount of manifold data concerning metabolic networks continually arising from current research activities in biotechnology leads to the great challenge of information visualization. Visualizing information in networks first of all requires appropriate network diagrams. In the context of metabolic networks, historical conventions regarding the network layout have been established. These layouts are not realizable by prevailing algorithms for automatic graph drawing. Hence, manual graph drawing is the predominating way to set up metabolic network diagrams. This is very time-consuming without software support, especially considering large networks with more than 500 nodes. We present a semi-automatic approach to drawing networks which relies on manual editing supported by two concepts of the interactive and automatic arrangement of nodes and edges. The first concept, called the *layout pattern*, uses an arbitrarily shaped skeleton as a backbone for the arrangement of nodes and edges. The second concept allows us to wrap a set of repeating drawing steps onto a so-called *motif stamp*, which can be appended to other parts of a diagram during the drawing process. Finally, a case study demonstrates that both semi-automatic drawing techniques diminish the time to be devoted for the manual network drawing process.

Keywords

biochemical networks, semi-automatic network layout, information visualization, graph drawing, metabolic pathways

Introduction

Information visualization has attained increasing importance in the field of systems biology and metabolic engineering where a wide variety of large-scale data arises continually from high-throughput experimental studies and computer simulations. Intracellular biochemical processes are usually considered networks whose nodes represent the actors of the complex interlaced biological system (e.g. metabolite pools, enzymes, genes) and the edges represent relations between these actors (e.g. conversion, activation, inhibition, degradation, synthesis).^{1,2} Networks describing the metabolic activities of a cell are called *metabolic networks*. They consist of enzymatic reactions transforming biochemical compounds into each other. These reactions compose a complex network of interdependencies and regulatory mechanisms.³ In metabolic networks, reactions can be classified by the so-called *metabolic pathways* being series of reactions that share a certain functional role in the metabolism.

Prominent pathways are, for instance, the glycolysis for breakdown of glucose and other sugars and the citric acid cycle producing high-energy reduction equivalents.

Visualization is a key to achieving an overall view of data corresponding to the components of a network. By this, deeper insights into complex interrelations between these network components can be achieved. A convenient, network-integrated visualization technique is to map scalar data to visual properties of network nodes and edges.^{3–7} In this way, a direct visual link is established between data and the

¹Institute of Bio- and Geosciences 1: Biotechnology, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

Corresponding author:

Katharina Nöh, Institute of Bio- and Geosciences 1: Biotechnology, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany.
Email: k.noeh@fz-juelich.de

biochemical network. It is important to visualize entire metabolic networks in a single drawing in order to identify interconnections on metabolic processes.

Layout conventions

There are several different types of biochemical networks from the genome, transcriptome, or proteome layer displaying different aspects of intracellular processes.⁸ Each of these layers has its own, very specific peculiarities. The network type we focus on is the metabolic layer. In the context of metabolic networks, we are concerned with historically established conventions for network layout. Biologists have these ‘familiar’ layouts in mind when they draw networks inspired by popular biochemical text books.^{2,9} These text books share *de facto* layout conventions that are commonly accepted but cannot be put into a list of fixed rules. However, there exist informal conventions.

The layout conventions concern the arrangement and shape of metabolic pathways in a diagram. For example, one significant motif, usually acting as landmark in comprehensive biochemical network drawings, is the circular arrangement of the so-called citric acid cycle (cf. Figure 2(d)). Even for organisms where the path is not closed (e.g. *Gluconobacter oxydans*), it is arranged in a sickle-shaped arc. Another example for a circularly arranged pathway is the so-called urea cycle. In contrast, the glycolysis (glyconeogenesis) pathway is usually arranged linearly from top to bottom (bottom to top) or from left to right (right to left). The pentose phosphate pathway and the anaplerotic section are usually arranged to the left or right to the glycolysis. The glyoxylate bypass is located inside of the citric acid cycle. The amino acid synthesis pathways are usually drawn to both sides of the glycolysis and citric acid cycle in a more or less linear scheme. These conventions are informal and, by this, imply certain flexibility. The most important problem for the computation of these layout conventions is the actual composition of the pathways is organism-specific and varies between different biological systems. Despite the fact that metabolic networks are widely utilized, there is still no common standard that unifies the identifiers of metabolite pools and reactions.

When simulated and experimental data have to be visualized in a network context, the established layout conventions should be obeyed in order to increase the acceptance and readability. Following established layout conventions facilitates rapid orientation. The human observer can concentrate on the essential quantitative or structural information related to the network. This makes discussion of, for instance, measurement values visualized in a metabolic network diagram much more efficient.

Automatic versus manual approaches

As an illustrative example, Figure 1 shows a small network of a microbial central metabolism. In Figure 1(a), the network was manually drawn according to the standard text book layout. Figure 1(b–d) display the same network but are designed by means of three different automatic layout algorithms. Life scientists and biologists are immediately able to identify all the essential metabolic pathways shown in Figure 1(a). It is quite obvious, however, that the automatic computed diagrams are hard to read because even the prominent paths and cycles cannot be recognized in the automatically computed arrangement (Figure 1(b–d)). These diagrams are inadequate for information visualization where orientation and fast identifiability are indispensable prerequisite as it is the case for metabolic networks.

As stated by Suderman and Hallett, most tools in the context of biochemical network visualization support a fully automatic network drawing mode.¹⁰ Here, the usual approaches of automatic graph drawing, as for instance hierarchical¹¹ or force-directed^{12,13} methods, are adapted and improved according to the requirements of the biochemical application field by, for example, clustering or grid-based^{14–16} approaches. However, Saraiya et al. wrote that the utility and impact of those approaches for visualization of metabolic networks is yet unclear.¹⁷ Even these adapted automatic solutions rarely lead to well-accepted network layouts.

This contribution does not argue against automatic graph drawing in general. Certainly, these techniques are successfully applied in other types of biochemical networks, for instance, protein-protein interaction networks.¹⁸ In such application fields, the arrangement of nodes and edges in the diagram has no effect on the identifiability of single components because there are no standardized layout rules. Hence, an automatic layout of network components does not affect the gain of insight from the visualization.

Automatic graph drawing is well suited if no commonly accepted layout conventions exist, or, graphs are generated for getting a qualitative impression by means of connection-centric approaches. Furthermore, automatic network layout is undoubtedly valuable where the size of the networks exceeds several ten thousands of nodes even in the context of metabolic networks.^{19,20} Gehlenborg et al. state that although the available automated layout methods may be useful, they usually lead to low-quality layouts compared to manually drawn networks and often require manual editing in addition.²¹ Human intervention is necessary in order to adapt the automatically computed layout to the requirements of the scientific application field.²² Here, we focus on smaller

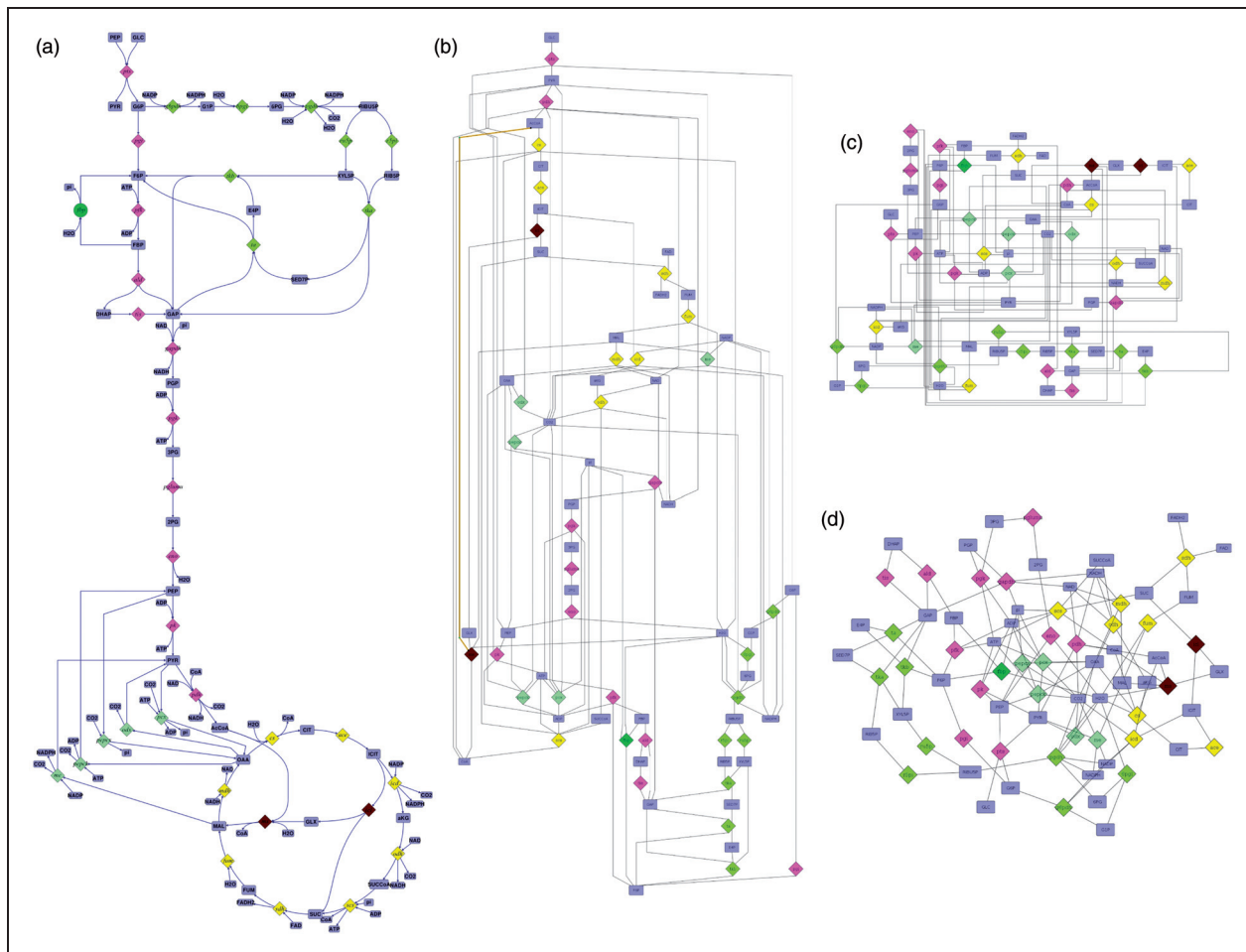


Figure 1. Layout of the central metabolic pathways: (a) manually drawn according to well-established layout conventions; (b) to (d) automatic computed graph layouts with several algorithms provided by the graph editor yEd (<http://www.yworks.com>).

networks ranging from one hundred up to two thousand nodes.

Circular arrangements

Prevailing automatic graph layout algorithms are not able to arrange cycles as they are commonly drawn in metabolic networks. From a graph theoretical point of view, metabolic networks usually contain several potentially nested cycles as depicted in Figure 2(a–d). Here, a section of Figure 1(a) is shown with four exemplary graph cycles highlighted in the network. Not all graph cycles, however, are actually drawn as circles in a network diagram by convention. In the example from the central metabolism, only the citric acid cycle (cf. Figure 2(d)) is arranged in a circle as it was initiated by the pioneers of molecular biology. Obviously, graph analysis algorithms are not able to

identify the one network path conventionally drawn as circle.

A frequently applied automatic layout scheme is the arrangement of all network nodes on a circumference interconnected by straight line edges²³ as shown in Figure 2(e)). However, a circular arrangement of cyclic network paths as part of a hierarchical network layout is rare and to our knowledge has not yet been applied successfully.

Layout requirements of metabolic networks

We argue that the design process of metabolic network diagrams is governed by conserving *recognition value*, which means rapid orientation in the diagram and immediate identifiability of ‘familiar’ network sections. Recognition value is very important in order to generate network-integrated data visualizations giving fast impressions of the embedded information.

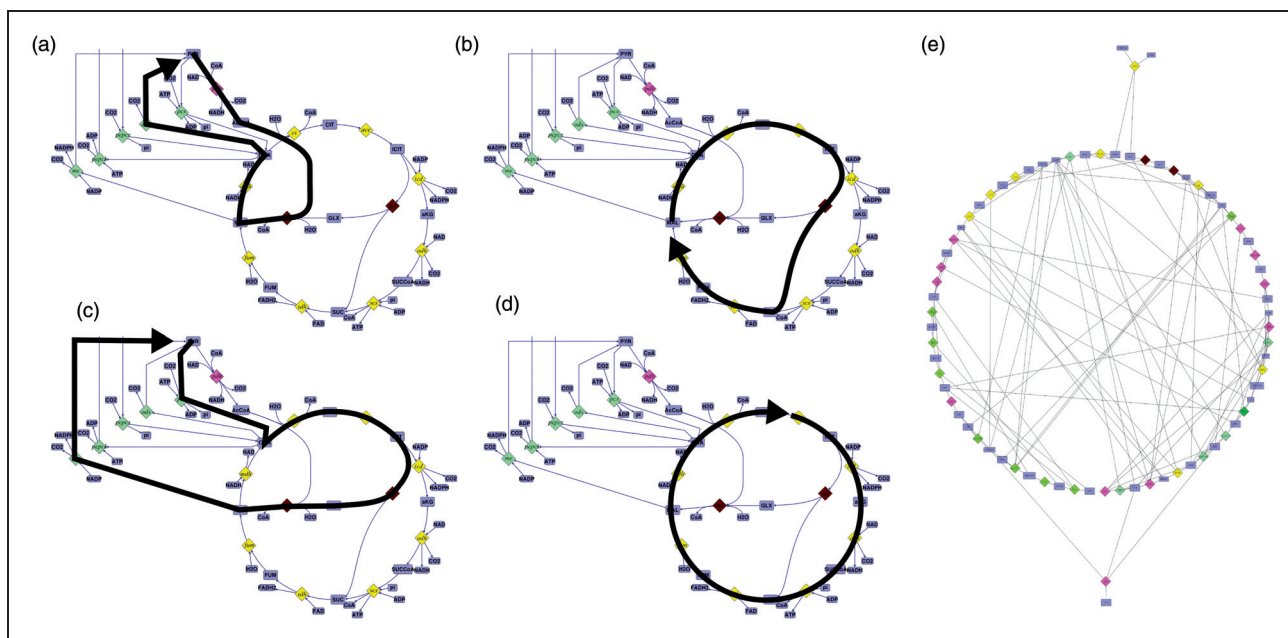


Figure 2. Circular arrangement of paths within nested metabolic networks: (a) to (d) examples for cyclic paths in one subnetwork of (e) circular network layout automatically generated by yEd.

Another closely related aspect in drawing metabolic networks is *aesthetics*. Here, aesthetics does not only and not necessarily include a minimized number of edge crossings and bends, symmetry, and maximal angles between edges leaving a node.²⁴ According to Saraiya et al., the applicability of existing concepts on network aesthetics in the context of metabolic networks is not clear.¹⁷ We propagate a more general concept of the term aesthetics including the shape of edges and proportions of nodes. Figure 1(a) shows differently formed edge shapes leading to an aesthetically pleasing diagram. Here, edges between circular-arranged nodes are shaped as arcs. Other edges have linear or curved characteristics adapted to the arrangement of nodes. A third aspect that becomes increasingly important for larger networks is *efficiency*. Drawing networks is very time-consuming. Saving time by supporting the human activities is a desirable goal for drawing software for metabolic networks, e.g. by automating repetitive actions.

These three requirements – recognition value, efficiency, and aesthetics – must be supported by a network drawing tool. We argue that it is nearly impossible to perform all these requirements for metabolic networks with automatic layout algorithms. Clearly, the automatic placement of network elements is time-efficient, but the designed network diagrams do neither satisfy common aesthetics nor have any recognition value (see Figure 1). In the context of metabolic networks, we demonstrate that manual graph drawing

supported by a small set of automatisms leads to the widely accepted network layouts whereas the time effort for the drawing process is diminished.

Contribution

We present a novel semi-automatic graph drawing approach that satisfies the special requirements of metabolic networks. This approach consists of the so-called *layout pattern*, which is described in Section ‘Layout pattern: shaping the coarse network structure,’ and the *motif stamps* (see Section ‘Motif stamps: cloning small reaction patterns’). Section ‘Work Flow’ introduces a work flow for applying the design and layout techniques. In order to give an impression of the strength of our contribution Section ‘Case study’ discusses a case study in ‘real life’. First of all, Section 2 gives a short introduction to the terminology of metabolic networks on which this work is based.

Metabolic networks

Metabolic networks are commonly represented as directed hypergraphs^{7,25} where metabolites are symbolized by nodes and the connecting edges correspond to the enzymatic conversions (reactions). Here, every edge may have several sources and destinations (i.e. substrates and products of a reaction step). However, such directed hyperedges can also be represented in a

bipartite graph structure by introducing individual nodes for reactions.^{15,22} Our approach relies on the latter convention (cf. Figure 3).

Graph definition

The above-described network concept leads to a directed graph $G=(V, E)$ describing a metabolic network with set of nodes $V=M\cup R$ and $M\cap R=\emptyset$ containing metabolite nodes M and reaction nodes R connected by edges $E\subseteq\{M\times R,R\times M\}$. A directed edge $e\in E$ between a start node v_1 and a destination node v_2 , $v_1, v_2\in V$ is denoted by $e=(v_1, v_2)$. A set of all edges connected to a node $v_1\in V$ is denoted by $E(v)\subseteq E$ and the connectivity of v by $|E(v)|$. The set E can be subdivided into disjoint types of edges used in metabolic networks. The set of ‘flux edges’ $f\in F$ representing flux relations between metabolites and reactions and the set of ‘effector edges’ $x\in X$ indicating regulatory effects between metabolites and reactions³ are given by:

$$F = \{(v_1, v_2) | v_1 \in M, v_2 \in R \vee v_1 \in R, v_2 \in M\} \quad (1)$$

$$X = \{(v_1, v_2) | v_1 \in M, v_2 \in R\} \quad (2)$$

Metabolic pathways

Classification of reaction sequences into metabolic pathways provides important meta-information

about the network structure. It can help to compute network layouts because these pathways are usually drawn as a linear or circular node sequence. A metabolic pathway $P\subseteq R$ is a set of enzymatic reactions defining a special functional module of the network. Let P_1 and P_2 be two pathways, then $P_1\cap P_2=\emptyset$. Like metabolites and reactions, pathways are also labeled with a unique identifier. The periphery $\Pi_r=(E_r, M_r)$ of a reaction r is given by its connected edges $E_r=E(r)$ and the connected metabolites $M_r=\{m\in M | \exists e\in E(m)\cap E_r\}$ (distance 1). The peripheries of all reactions of a pathway P lead to the ‘pathway subgraph’ which we denote $G_p=(P\cup\bar{M}, \bar{E})$ with:

$$\bar{M} = \bigcup_{r\in P} M_r, \quad \text{and} \quad (3)$$

$$\bar{E} = \bigcup_{r\in P} E_r \quad \text{with} \quad \bar{E}\cap X = \emptyset \quad (4)$$

Note that effector edges are excluded from the set of pathway edges.

Duplicated nodes

A distinguishing feature of metabolic networks is the existence of a few nodes that show a high degree of connectivity. This especially applies to certain metabolites like ATP or NADH, which participate in many enzymatic reactions over the entire network.

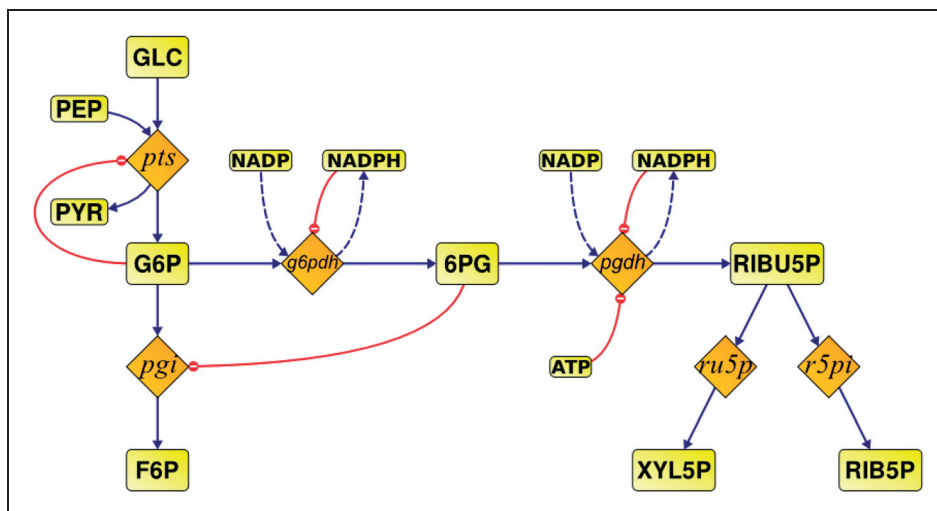


Figure 3. Bipartite graph representation of a metabolic network example. Reactions are displayed by diamonds, metabolites by rounded rectangles. The edges with a triangular arrowhead indicate a flux relation. The edges with a circle end marker indicate a regulatory effect. The metabolites NADP and NADPH are represented by two duplicated nodes in each case.

Displaying a metabolic network with only one representative of each of those highly connected nodes would lead to a chaotic diagram with a myriad of edge crossings. Avoiding these edge crossings leads again to uncommon network layouts. In order to reduce edge crossings without affecting readability, we introduce the possibility of duplicatable metabolite nodes. By this, a metabolite can appear several times in a network diagram; or in other words, several metabolite nodes can represent one single metabolite $\hat{m} = \{m_n | m \in M, n \in N\}$. In this way, the connectivity of the visual representation is strongly reducible.

Creating network diagrams

As mentioned above, this work introduces novel automatism that support and accelerate the process of manual graph drawing. These techniques are implemented in the network drawing tool ‘Omix.’ Omix is a software solution for customizable information visualization in the context of metabolic networks. With Omix, metabolic network diagrams can be created and data from experimental studies or simulation can be visualized in these diagrams as published elsewhere.^{6,26}

Figure 4 shows the main window of Omix consisting of a drawing area in the center where diagrams are drawn, several toolbars giving access to functionalities of the software and two sidebars showing diverse information about the network and its components in structured manner.

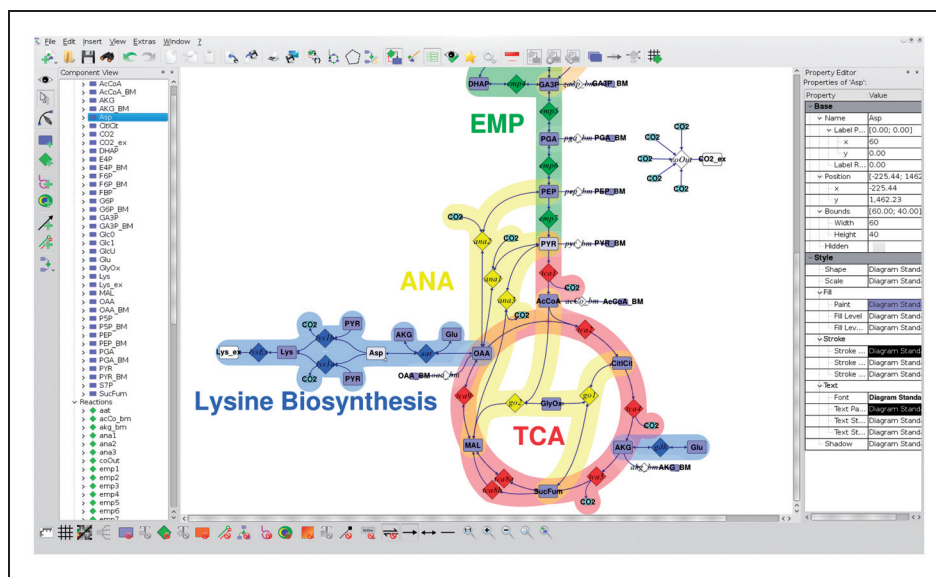


Figure 4. Main window of the network drawing tool Omix among others consisting of a drawing area (center) and a component list sidebar (left).

Basically, there are two alternatives for setting up network diagrams in Omix:

- A network diagram can be manually drawn bottom up, i.e. every node is drawn individually and every edge is inserted one by one.
- A network diagram can be generated by importing a network from third party file formats or from databases. Here, usually the network components' layout information is not available. Hence, the nodes and edges must be arranged in the diagram.

Both options are quite time-consuming without any software assistance even for networks with a moderate number of nodes. The following sections introduce techniques realized in Omix in order to speed up the drawing process.

Layout pattern: shaping the coarse network structure

The drawing area of the diagram editor can be pre-structured with a so-called *layout pattern*. A layout pattern is a skeleton on which series of nodes and edges can be arranged (see Figure 5). In the process of arranging nodes on the drawing area, the layout pattern supports the user in finding the correct node sequences in a heterogeneously connected graph. To a certain extent, the layout pattern is a meta-component of the diagram imprinting an overall structure to the network. In the following, this novel method of semi-automatic network layout is

introduced in more detail. The usage of the layout pattern for node and edge arrangement is illustrated and the particular challenges arising from the combination of interaction and automation are stressed.

The layout pattern is an arbitrarily complex figure consisting of several spline segments, the so-called *pattern sections*, which can be versatily assembled. Figure 5(a) shows a pattern figure for the central metabolism which was already illustrated in Figure 1. Every pattern section is an arbitrarily shapeable spline between two branch points as illustrated in Figure 5(b). On a pattern section, a sequence of nodes and edges is arranged according to the geometry of the path curve (see Figure 5(c)). When the shape of a pattern section is subsequently changed, the position of all its nodes and the shape of all the edges are immediately adapted to the new shape of the pattern. A node on a pattern section can have two neighbors; a node on a branch point can have as many neighbors as pattern sections are connected to the point.

Pattern sections can be added to the drawing area. Here, user guidance is oriented to the curve handling of common graphics software tools. Curvature can be edited with Bézier control points, smoothness can be enforced, and curve sections can be joined or split. In this way, the favored diagram layout can be prepared.

Network editing and layout pattern editing work in two different layers. When the pattern is edited, it

becomes a top-level item, whereas all components of the network diagram are inaccessible and semi-transparently displayed in the background (cf. Figure 5(b)). Vice versa, the layout pattern is semi-transparent in the background in network editing mode (Figure 5(c)). When the diagram is used for visualization, printout etc., the layout pattern is invisible.

Using the layout pattern

This subsection gives an impression of the work flow of using the layout pattern. As mentioned above, nodes and edges can be placed on the layout pattern. This is exemplified in the image series in Figure 6, where a chaotic network is being arranged by a user-defined pattern section (see Figure 6(a)). Using the layout pattern is done by positioning nodes either on top of a pattern section or a branch point. In Figure 6(b), a reaction node has been moved onto a circular pattern section. When a node $v_0 \in V$ is added to the layout pattern in this way, the layout pattern searches for paths in the network that interconnect the added node v_0 to its neighboring nodes $v_0 \dots v_n$, $n > 1$. The path search requires human interaction because there can be many paths between two nodes in a network. The challenges and requirements of this path search are discussed below.

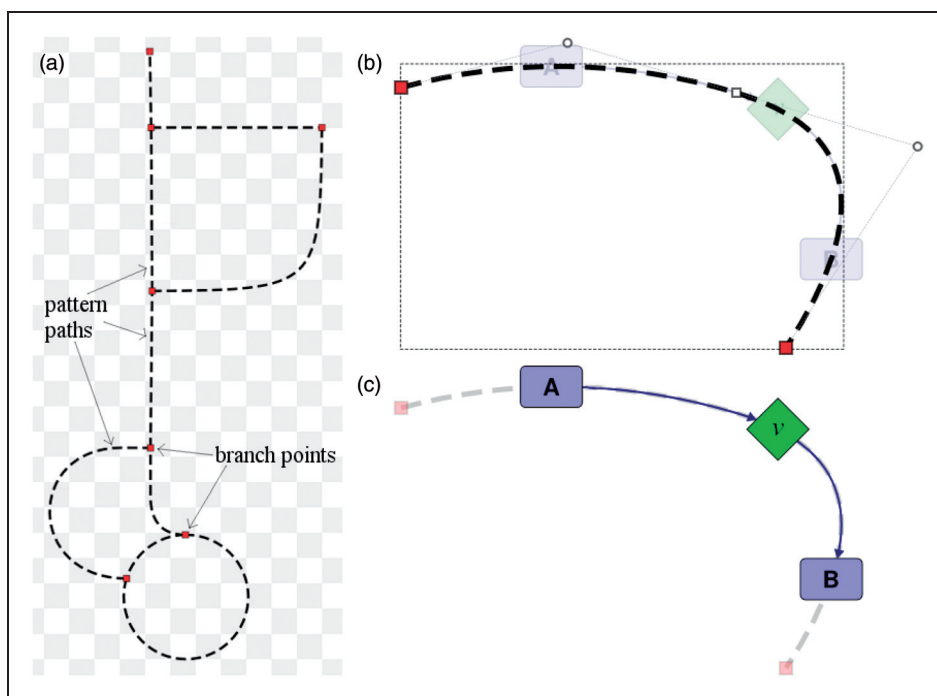


Figure 5. Layout pattern consisting of several pattern sections connected by branch points (a). Pattern sections formed as sequence of Bézier curves (b). Arrangement of nodes and edges on a pattern section (c).

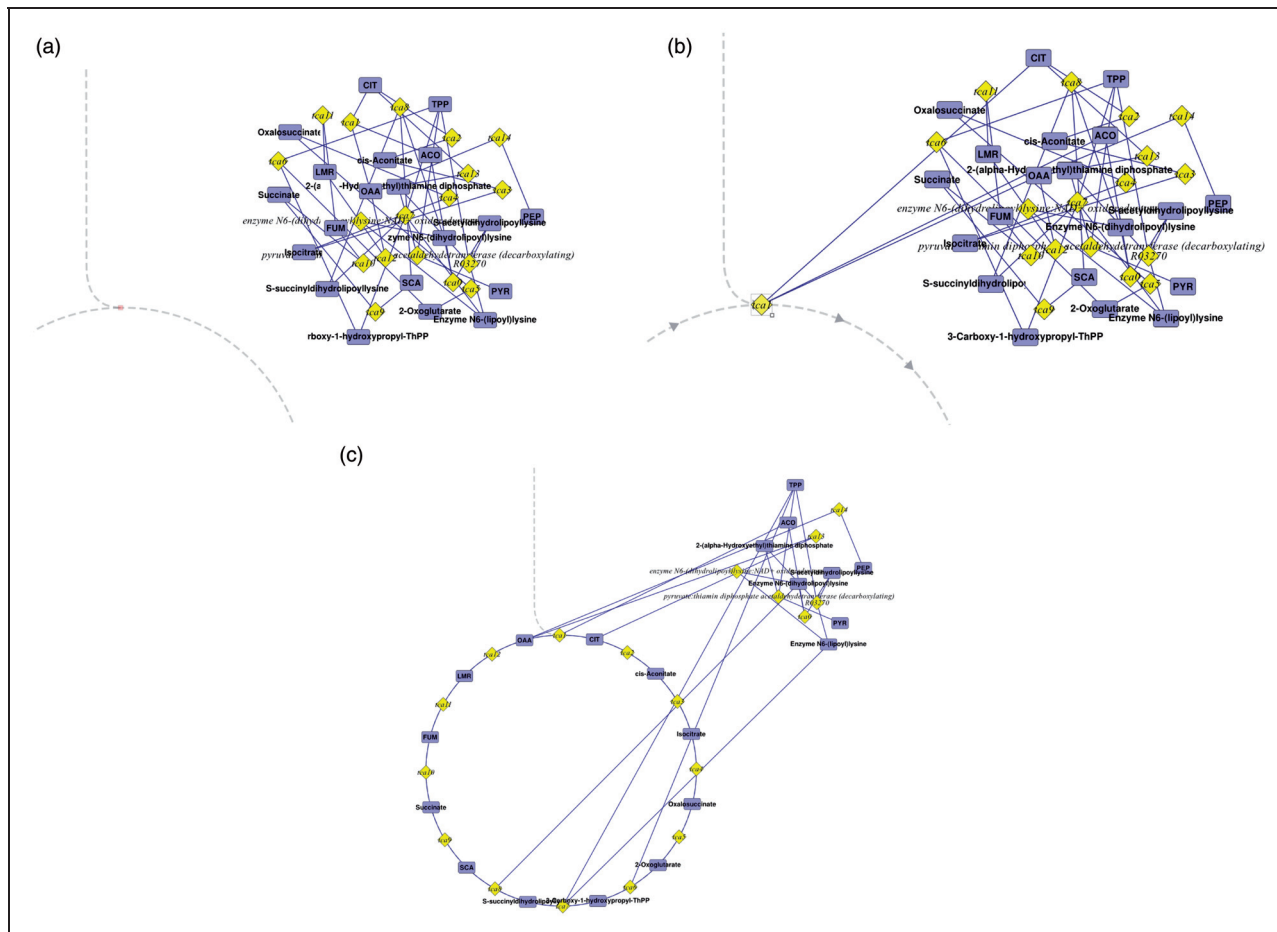


Figure 6. Using the layout pattern to design a sequence of nodes and edges: (a) initial situation, (b) starting the layout step by moving a node on top of the pattern, (c) finished layout step after interactive search of matched node sequence.

Since the pattern section in the example shown in Figure 6 is a circle, a cyclic path must be specified by the user to be arranged on the pattern. This path starts and ends with the reaction node recently added to the layout pattern. After selecting the preferred path, its node sequence is added to the pattern by arranging the nodes equidistantly while reshaping the edges according to the form of the underlying pattern section (cf. Figure 6(c)).

Subsequently, further pattern paths can be inserted for the placement of the other node sequences in the network. Thus, the layout pattern can be used to successively arrange metabolic pathways in the diagram.

Challenges and requirements

Human interaction is required to decide which network path is to be arranged between two nodes v_0 , v_n in the network positioned on the layout pattern. The user has to select the preferred path out of all

possible network paths. A prior path search on the graph is necessary in order to compute all possibilities. Here, the following conditions have to be taken into account:

1. The directed metabolic network has to be regarded as an undirected graph because the pattern sections are non-directional. This convention enables the user to arrange sequences of opposed edges on the layout pattern.
2. A node may not occur twice in a network path except for it is a start as well as an end node on a circular pattern section.
3. Only those nodes that are not already positioned on a pattern section may occur in a network path.

The number of network paths between two nodes is $O(n!)$, with $n=|V|$ in the worst case of a complete graph. Considering a maximal number k of edges connected to a node, the complexity is diminished to

$O(n^k)$ which nevertheless leads to two non-trivial problems:

1. Calculation of all possible paths with standard search algorithms like depth or breadth search leads to computation times, which are unacceptable for human interaction even for small networks.
2. Selecting the preferred path from an exhaustive list of all possible ones typically requires distinguishing between a huge number of candidates. This is nearly impossible even for experienced users.

Due to both of these reasons, fully automatic network analysis is unsuitable in our context. Constraining the analysis, for example, by a limited depth search certainly reduces the computation time. However, a limited depth search leads to an incomplete set of result paths. Due to these runtime and completeness problems, we developed a novel concept of network path search to assist the user in selecting the preferred path in an economic and interactive way.

Semi-automatic path search

In order to deal with the mentioned runtime and completeness problem, the node path search of the layout pattern is implemented in a semi-automatic manner. Here, the core idea is a user-performed manual depth search which is software-supported by a previewing and preselecting search algorithm.

Human-interactive path search

The human interaction interface is a tree view (see Figure 7) representing the graph in the form of a tree whose root element is the node v_0 currently moved above the layout pattern. All nodes connected to the node v_0 are child elements of the node's tree element. The direction and type of an edge between two nodes is displayed by an arrow symbol preceding the node name as shown in Figure 7.

Initially, only the root element and its first nesting level elements are visible. The user has to traverse the tree by expanding its elements until the target node v_n

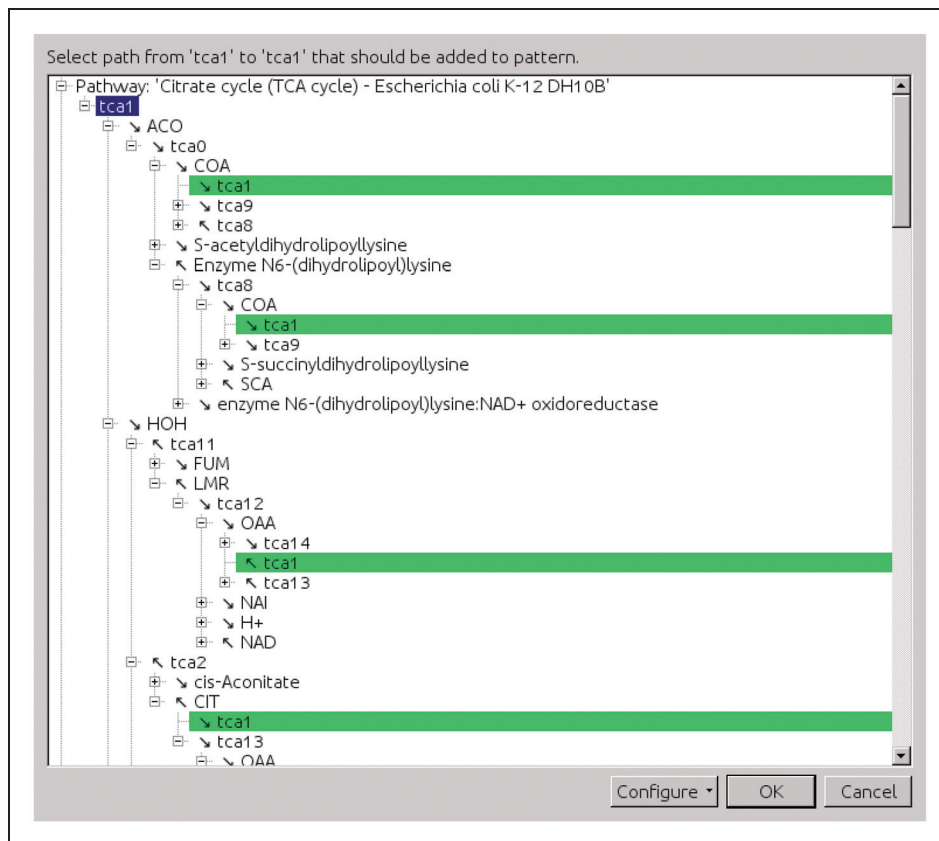


Figure 7. Dialog window of the interactive node path search. In the scenario shown, the user is asked to select a cyclic node path containing the node "tca1." The internal search algorithm has already determined several valid node paths. These paths are expanded and highlighted in the tree. The arrow symbols indicate the direction of the interconnecting edges.

occurs as a leaf element in the tree. A leaf element containing v_n represents a valid path which can be added to the layout pattern. For a fast perception, the target leaf elements are highlighted in green as pictured in Figure 7. By selecting a leaf element, the corresponding node path is designated for automatic arrangement on the layout pattern.

Software-assisted preselection and exclusion of network paths. When the tree is initialized and, likewise, every time the user expands a new tree element a search algorithm traverses the graph. Triggered by expanding a tree element, the algorithm is initialized with the currently expanded element. Thus, the software always pre-caches the network topology several nesting levels in advance of the user's manual search activity. Here, a depth-limited search algorithm is applied with an initial recursion limit $d = 15$, whereas the limit is decreased in each recursion step depending on the connectivity $|E(v_i)|$ of the currently traversed node v_i :

$$d_{next} = d_{current} - |E(v_i)| \quad \text{with} \quad v_i \in V \quad (5)$$

In particular, in the presence of highly connected nodes, the context-dependent adjustment of recursion depth avoids the impairment of the human-computer interaction.

The software search is conducive to the simplification and acceleration of the user search. A valid network path ends with the target node v_n . If v_n is found during the search procedure, the complete branch path leading to the node is automatically expanded. The user can inspect and optionally select the complete path immediately.

When a node occurs twice in a network path, it is invalid. If the depth search algorithm detects an invalid path, the corresponding branch is eliminated from the tree. If an eliminated branch was the only child element of its parent, the parent element is also eliminated. This continues until a parent element has at least one child element left. If the depth search does not lead to an assertion about the validity of a path, the corresponding tree element remains unexpanded in the tree and may be expanded manually by the user on demand.

Eliminating elements representing invalid paths prevents the user from traversing dead branches and from selecting paths that contain loops. If no path exists between two nodes in the network, the software search will successively traverse the complete graph and eliminate all elements in the tree view. In this case, no further nodes and edges are arranged on the layout pattern.

In case of short network paths between the nodes v_0 and v_n , the software search can find the preferred path immediately during the initial traversing and offer it for user selection. Furthermore, the user can successively trigger the complete analysis of the network leading to the elimination of all invalid paths in the tree view.

Pathway-oriented layout. As mentioned in Section 'Metabolic pathways,' metabolic pathways lead to a biologically motivated breakdown of the main graph G into smaller subgraphs G_p . These pathway subgraphs consist usually, but not exclusively, of linear or circular node sequences surrounded by short branches.

If both nodes v_0 and v_n added to the layout pattern are part of the same pathway subgraph G_p , the interactive node path search is optionally performed on G_p in addition to the search on the complete graph G . This, additionally, accelerates the interactive path search process.

Motif stamps: cloning small reaction patterns

Whereas the layout pattern concept helps to roughly arrange network diagrams, the second innovative technique of the semi-automatic graph layout is concerned with a more detailed view: the so-called *motif stamps* are intended to speed up the drawing of similar and multiply occurring compositions in a network diagram.

The concept of motif stamps aims at simplifying similar actions during the drawing process of a metabolic network. A motif, in this sense, is a template periphery Π_r (see Section 'Metabolic pathways,') surrounding a pseudo-reaction, which can be appended ('stamped') to other reactions in the network diagram. If a motif is appended to a reaction, all metabolites and edges are inserted according to the motif. The term 'motif stamp' is derived from similar concepts in pixel-oriented graphics software tools.

The concept of motif stamps is motivated by the observation that certain metabolic reactions are similarly composed and connected metabolites and edges are frequently arranged in a similar manner. Figure 8 shows a subsection of a large network diagram. The cofactors ATP and ADP are involved in many reactions all over the network as highlighted in the figure. Many other similar cofactor pairs are frequently involved all over a metabolic network like NADP and NADPH, FAD, and FADH. Other frequently occurring reaction partners are for example H_2O , CO_2 , and ions. It is a well-established practice to display these highly connected metabolites by one node

duplication per occurrence with similar edge shapes (cf. Figure 3). This leads to many similar actions the user has to perform during the drawing process, which includes duplicating and arranging nodes, inserting edges, adapting to the shape and visual properties of the edges and so on.

Figure 9 shows a schematic diagram of the motif stamp concept. Metabolites and edges are arranged around a pseudo-reaction composing a motif stamp as shown in Figure 9(a). The pseudo-reaction is not part of the drawing but template for any reaction, the motif is later appended to a reaction (Figure 9(b)), the metabolites and edges are arranged around the reaction node as defined in the motif stamp (see Figure 9(c)).

Creating motif stamps

The motif can be designed according to individual aesthetic aspects. A motif stamp defines the connected metabolites of a reaction, their position relative to the

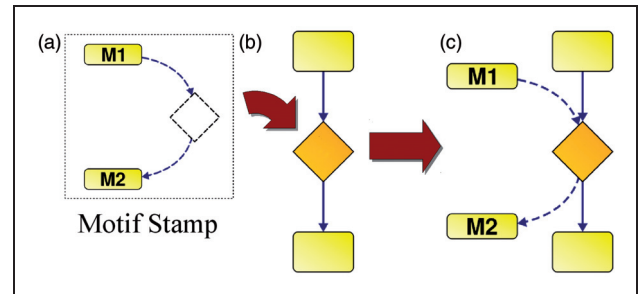


Figure 9. Schematic diagram of the use of motif stamps in a drawing process.

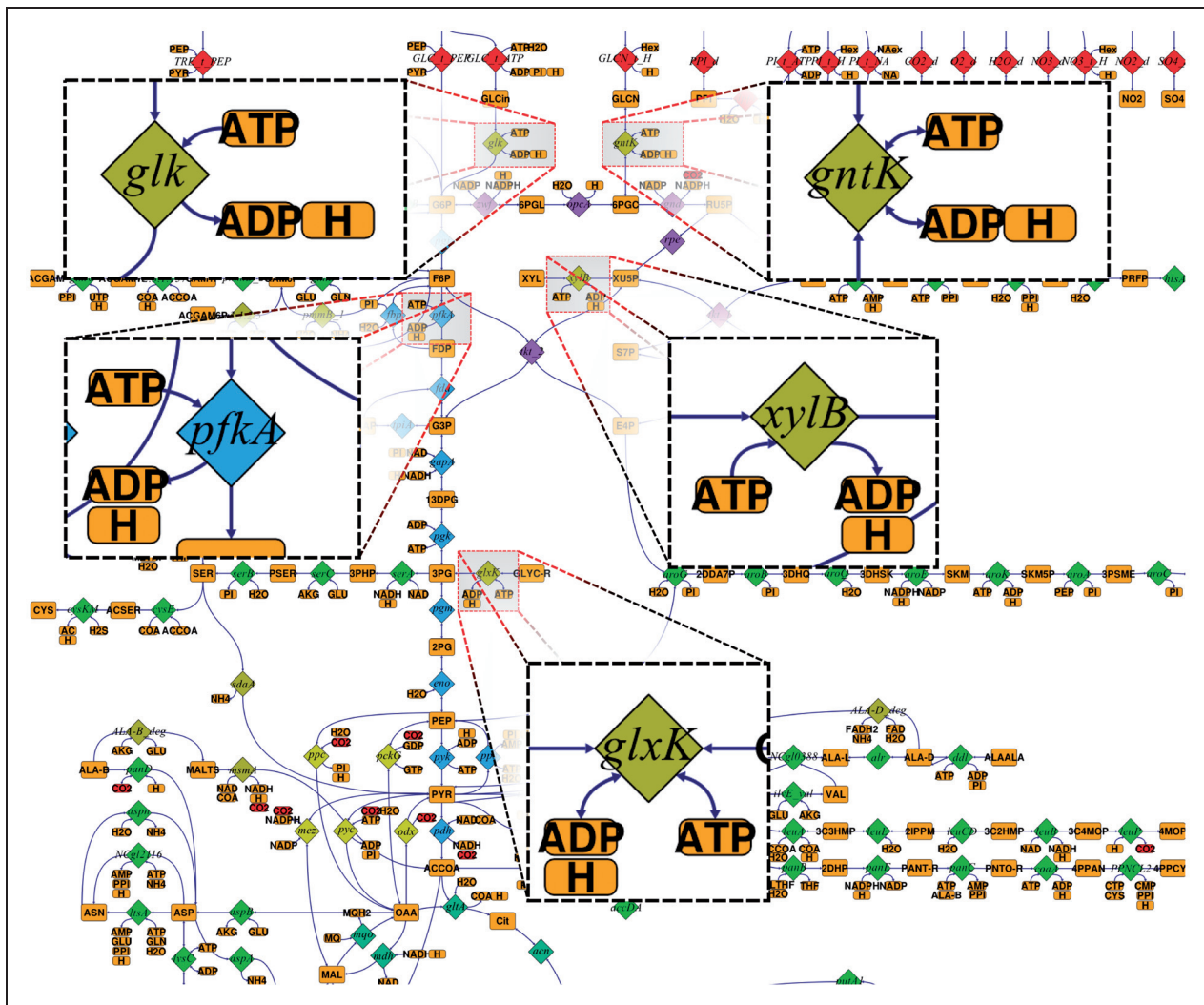


Figure 8. Multiple-occurring reaction partners ATP, ADP, and H.

reaction node and the shape of the connecting edges. Furthermore, visual properties like color, bounds, or text font of the nodes and edges are also inherent in a motif stamp.

Motif stamps can basically be created in two ways:

1. The motif can be drawn step by step. Here, metabolites and edges are created and connected to the pseudo-reaction node. Any number of nodes can be inserted, but there should be at least one edge between every node and the pseudo-reaction.
2. The composition of a motif can be captured from the periphery of an existing reaction in the network diagram. After selecting a reaction as a template for a new motif, the user has to specify which connected metabolites should be part of the motif stamp and which are dispensable components.

After naming the motif stamp, the definition process is finished. In this way, multiple motif stamps can be defined containing certain constellations and graphical appearances of reaction partners.

Appending motifs to reactions

Motifs can be appended in three steps: selecting a preferred motif stamp, selecting a reaction for appending the motif, and finally adjust the motif's orientation. A motif stamp can be rotated, scaled, and inverted in order to adjust it to the environment of the reaction it is

appended to. These transformations are illustrated in Figure 10 showing the semi-transparent silhouette of the appended motif around a target reaction node. The user can define the angle for the placement of the motif (Figure 10(a)). Figure 10(b)) shows a scaled version of the motif stamp. In Figure 10(c), the motif is mirrored and aligned with the reaction's connections. After confirming the alignment angle, the motif is finally appended to the selected reaction node (Figure 10(d)).

In order to aid aesthetics, the software optionally helps to align the motif stamp in a smooth manner with respect to the already existing periphery of the reaction. Therefore, an axis is computed for the appended reaction node. The axis is the average of the tangents of all edges at the docking position of a reaction. Likewise, motif stamps have an axis which is indicated by a hairline (see Figure 10(a–c)). During the motif rotation by the mouse, the user can choose to automatically align the motif's axis to the axis of the reaction node (cf. Figure 10(c)).

When a motif is appended to a reaction, three different cases can occur and have to be taken into account (cf. Figure 11):

- a. The target reaction is not yet connected to a metabolite M defined in the motif stamp – In this case, the metabolite M is inserted and connected according to the motif stamp.
- b. The target reaction is connected to the metabolite M and the metabolite node has further

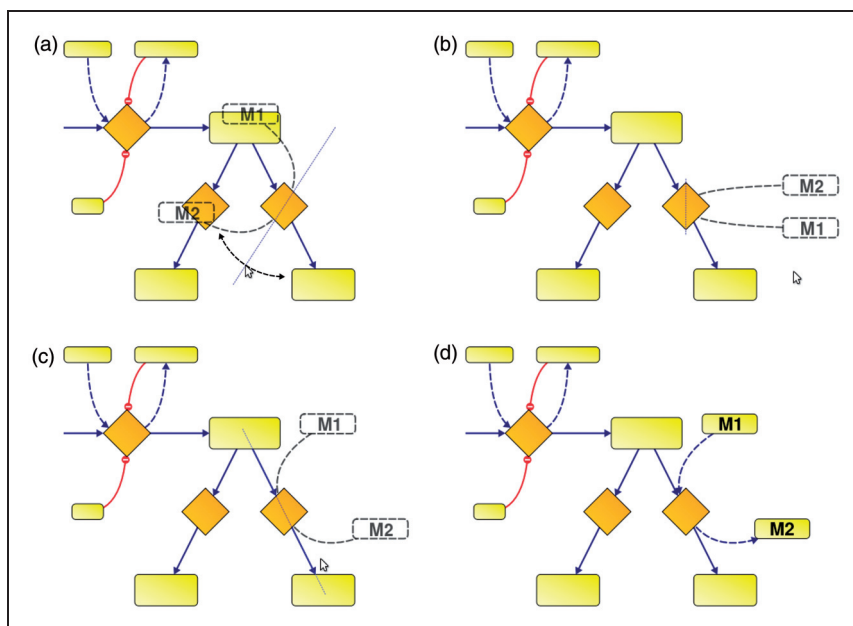


Figure 10. Operations to append a motif stamp to a reaction: (a) rotation, (b) scaling, (c) automatic “snapping” to a smooth arrangement, (d) final result.

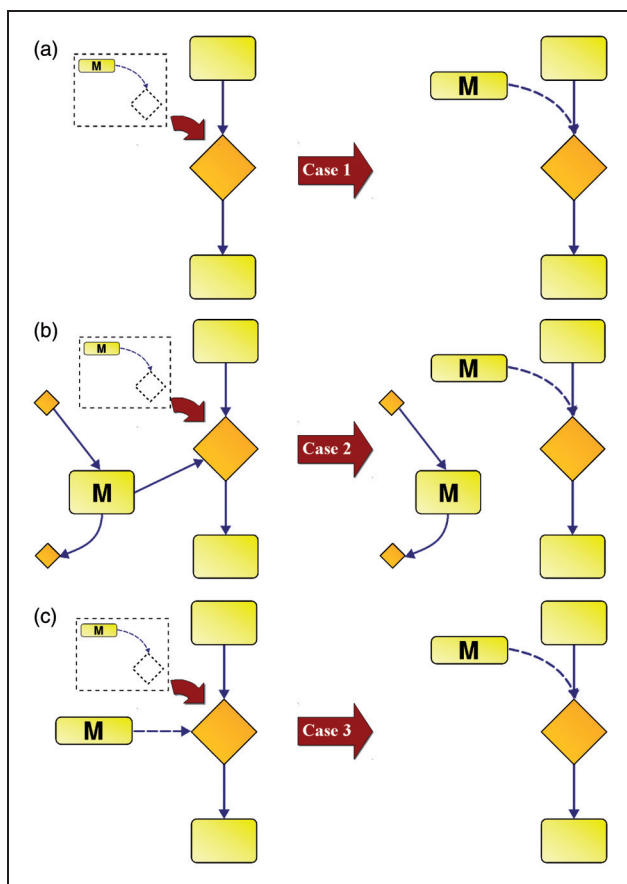


Figure 11. The motif stamp defines the connection between a metabolite M and a reaction including curve shape, node size, and other visual properties. When a motif stamp (dashed box) is to be appended to a reaction, three cases arise differing in whether and how M is already connected to the appended reaction: (a) not connected, (b) connected among other connections, (c) solely connected.

connections – Here, a further duplicate of the metabolite M is inserted. The connecting edge is switched to the new copy and arranged as defined in the motif stamp. If the metabolite node does not yet have any layout information (see Section ‘From layout-less nodes to network diagrams’), the node and its edge is automatically inserted into the network diagram.

- c. The target reaction is already connected to the metabolite M, which has no further connections – only the arrangement of the node and edge has to be adapted to. This also holds if the node has no layout information.

These three cases apply on every single metabolite connection defined in a motif stamp.

The motif stamp concept allows multiple motifs to be successively appended to a reaction node. Hence, a more complex constitution can be assembled from several motif stamps. The reaction periphery composed in this way can furthermore be a template for a new, larger motif stamp.

Summarizing, it has become clear that motif stamps can be used to initialize new connections as well as to arrange existing connections between one or more metabolites and a set of reactions in the diagram. The usage of motif stamps can accelerate the drawing process considerably because metabolites do not have to be cloned manually and edges do not have to be drawn and reshaped every time a certain repeating configuration is needed.

Work flow

This section describes a work flow of drawing a biochemical network diagram and, hereby, shows in which way the techniques of semi-automatic layout introduced above accelerate the drawing process. In this example, a metabolic network is given in the form of a model description file (e.g. in the Systems Biology Markup Language – SBML²⁷) without layout information which is to be drawn with the network drawing software Omix.

After importing the network, all reactions and metabolites appear in the component view from where they are added to the diagram by drag-and-drop (cf. Section ‘From layout-less nodes to network diagrams’). An important precondition for the semi-automatic network drawing solutions as presented here is knowledge of the semantics of the network topology on the user side.

Metabolic pathways

Before the network diagram is drawn, it is recommended that reactions should be grouped into metabolic pathways (cf. Section 2.2). The SBML format, for instance, does not directly support this kind of information; hence, it is not available after model import by default.

The user can add pathways to the network model and assign reactions to each of them. This optional task leads to the decomposition of the complex graph into smaller, less complex subgraphs.

From layout-less nodes to network diagrams

The main window contains a sidebar listing all components of the network (cf. Figure 4). When a network is imported without any layout information provided,

node positions and edge shapes are initially not available. These nodes and edges are not instantly visible in the diagram. Instead, the corresponding node entries in the component sidebar are labeled as to be non-positioned. The nodes represented in this way can be added to the diagram by drag-and-drop. Only nodes can be inserted in this way. Initially invisible edges are inserted automatically as soon as the source and destination nodes are added.

Using the layout pattern

The arrangement of nodes and edges in the diagram can now be performed by applying the layout pattern concept. It is recommended to start with one of the central parts in the network. This is a common practice in drawing metabolic networks. Here, the citric acid cycle (TCA) is chosen as a prominent example.

The next step is to draw a circular pattern section in the pattern editor (cf. Figure 5(a)), which will later carry the nodes of the TCA. After dragging a metabolite or reaction node from the component view and dropping it onto the circular pattern section, the user is asked to select the network path representing the TCA in the network. This can be performed fast because the semi-automatic path search uses the above-mentioned classification into pathways as structural information about the network and, hence, offers the correct network path immediately.

After arranging the first pathway, the pattern can be enlarged by further sections in order to later arrange directly connected pathways. In this way, a structural backbone of the network diagram is drawn step by step.

Using motif stamps

After creating the structural backbone of the diagram motif stamps are used to arrange cofactors and connecting edges in the environment of reactions already inserted. For this purpose, several motif stamps can be drawn or imported from other network files. As an example, the metabolite CO₂ may be product of several dozen reactions of the imported network. When the CO₂ node is inserted in the diagram, all edges appear automatically, connecting the new node to the reactions in which CO₂ is involved. This results in many edge crossings.

Hence, a new motif stamp is created to duplicate the CO₂ node. A metabolite node CO₂ is inserted into the motif stamp and connected to the pseudo-reaction. The connecting edge can be reshaped as preferred. After creation, the motif stamp can be appended to all reactions of the network, particularly to all reactions that are already connected to CO₂ (cf. Section 'Appending motifs to reactions'.2). Every CO₂

connection is visually designed according to the arrangement in the motif stamp. The isomorphic design of similar parts of the network leads to an aesthetically pleasing diagram. In this way, all cofactors and other secondary reactants can be inserted in the network diagram.

Completing the drawing process

Finally, all remaining network components that have not yet been inserted by the layout pattern or a motif stamp are inserted manually by drag-and-drop. After finishing the arrangement of all imported network components, the visual style of the network can be adapted to individual preferences and requirements.

Visualization

The resulting designed metabolic network diagram can subsequently be used for a wide variety of visualization purposes. For this, Omix is equipped with extensive data visualization features. In order to be adaptable to any requirements and preferences of the user, Omix is programmable by a scripting language called Omix Visualization Language (OVL).^{6,26} OVL allows fast and simple access to visual properties of nodes and edges such as color, font, shape, and line width just to mention a few. In this way, visualization becomes highly customizable because the user can implement interactive components, data input operations, and mapping of data to visual properties of all network components. The features and possibilities of the OVL scripting language are discussed in.^{6,26}

Case study

As stated in the Introduction, the visualization of metabolic networks requires the conservation of *recognition value*, which includes the arrangement of network parts according to established layout conventions. Furthermore, aesthetics plays an important role. These two criteria cannot be fulfilled by automatic graph drawing and require an active human steering. On the other hand, efficiency becomes an important criterion when the time needed for manually drawing a network graph is the crucial bottleneck. Both semi-automatic layout techniques introduced in this contribution aim at accelerating the manual drawing process by simplifying and automating human drawing steps.

In order to investigate the effectiveness of the layout pattern and motif stamps for network drawing, we performed a case study consisting of two phases. In the first phase, a layout-less network has to be drawn with Omix, either with or without the assistance of the

semi-automatic techniques. In a second phase, the resulting diagrams were rated by a jury.

Sixteen researchers with biology or biotechnology background from our and four collaborating scientific groups participated in phase one. One half of the participants was encouraged to use the semi-automatic layout techniques (group 1) whereas the other half was banned from using them (group 2). The network to be drawn was medium-sized (44 metabolites and reactions composed in five metabolic pathways, cf. Figure 12). The network topology did not contain duplicated metabolites, but a number of highly connected cofactor metabolites. The latter ones are usually represented by duplicated nodes. A sample solution was not made available. In order to introduce the drawing capabilities of the tool and using the layout pattern as well as the motif stamps, we provided video lessons. After a short training phase, the network was to be drawn and the time effort for the drawing exercise was measured. It was up to the participants to choose a personal trade-off between drawing fast and drawing a ‘nice’ diagram. After finishing, group 1 was asked to assess the ease-of-use of both semi-automatic layout techniques and their benefit for solving the exercise.

In the second phase, the drawing results were examined by 10 experienced life scientists. In addition to the manual drawings, a set of diagrams have been created

by applying automatic graph layout algorithms on the given network. Every jury member received 20 randomly selected drawings from the set of automatically, full-manually, and semi-automatically drawn diagrams. Each drawing was rated with respect to its recognition value and aesthetics.

The results of the surveys are shown in Tables 1 and 2. Figure 12 shows one example diagram from the automatically generated (Figure 12(a)), the manually drawn (Figure 12(b)), and the software-assisted drawn networks (Figure 12(c)) in each case. The manual drawing process has taken about 2 h in average per participant. Compared to automatic graph drawing, this time span appears to be very long. However, related to the costs of drawing such a diagram in a classical way, that is by applying a common graphics software tool, 2 h are acceptable. In average, the semi-automatic layout methods accelerated the drawing process for about half an hour that roughly is 20% of the average duration (see Table 1).

The expert group valued the recognition value as well as the aesthetics of the automatic computed diagrams very low. The recognition value of both, full and partly manually drawn networks, were comparably rated in the upper third. That is not surprising because the central pathways of all diagrams were arranged according traditional conventions (cf. Figure 1(a)). The aesthetics of the semi-automatic drawings is

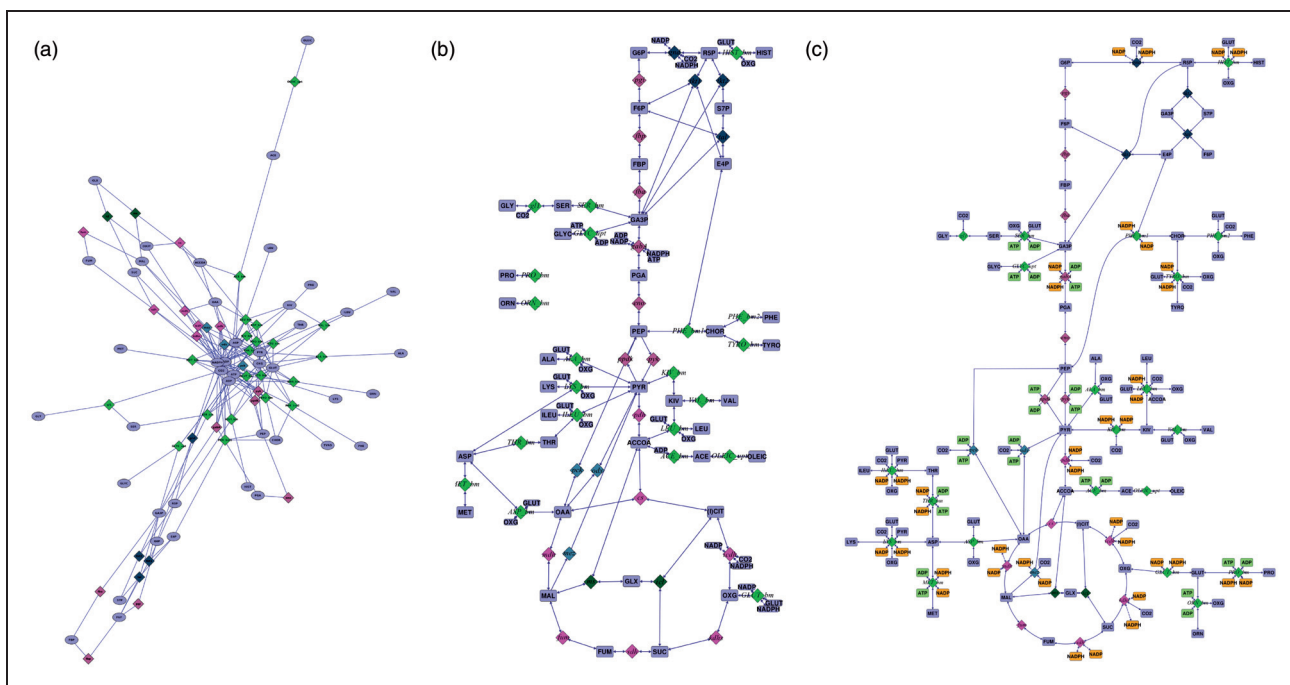


Figure 12. Selected results of the network drawing case study: (a) automatically (spring embedded layout), (b) manually, and (c) semi-automatically drawn network.

Table 1. Time effort in average per participant and ratings of the diagrams drawn with different approaches^a

	Time effort	Recognition value	Aesthetics
Automatic drawing	< min	1.6	1.7
Manual drawing	2:07 h	7.6	5.8
Semi-automatic Drawing	1:40 h	7.7	6.7

^aThe ratings were given from a scale between 1 and 10 (1 = very low, 10 = very high).

Table 2. Benefit of the two semi-automatic layout methods (1 = not useful, 10 = very useful) and ease-of-use (1 = very difficult, 10 = very easy)

	Benefit	Ease of use
Layout pattern	6.1	6.9
Motif stamps	7.0	6.4

slightly higher valued than the manual drawings. Here, it was noticeable that the nodes and edges of the TCA pathway in nearly all semi-automatic drawings were shaped to a perfect circle in contrast to the manual arrangements (cf. Figure 12(b and c)). This noticeably improves the aesthetic impression as well as the recognition value of a diagram.

As shown in Table 2, both semi-automatic layout techniques have been accepted by the users. The benefit for solving the exercise as well as the ease of use were assessed in the upper third.

The case study demonstrates that the semi-automatic layout techniques lower the time effort to manually layout a diagram according established layout standards while they simultaneously improve personal aesthetics. From further, informal discussions with the participants, we found out that the layout techniques were considered valuable for the arrangement of familiar pathways but that the most time-consuming issue was to understand unfamiliar parts of the network before being able to arrange them. This significantly caused the rather high time effort of about 2 h for drawing the network. In a scenario where the drawing process intimately lies in the research objective of the user, the layout techniques may be more beneficial. This scenario, however, was not intended by this case study and remains as an interesting future evaluation task.

The case study hardly represents an exhaustive evaluation since 16 participants cannot be regarded as a representative user pool. Another limitation is that only a single network has been tested. Nevertheless, the case study demonstrates or at least indicates the value of the presented semi-automatic layout approaches.

Discussion

Conclusion

This contribution presents a novel approach to semi-automatic graph drawing. In the application field of metabolic networks, special features have been established that require specific handling. Here, graph drawing is affected by historically evolved layout conventions. This greatly restricts the suitability of automatic graph drawing because recognition value is lacking in automatically arranged metabolic network drawings.

The present work introduces two concepts of software-supported manual graph drawing:

1. Layout pattern – a customarily shaped skeleton that arranges series of nodes and edges. This concept helps to design a layout-less network according to established conventions in a semi-automatic manner.
2. Motif stamps – wrapping a number of drawing steps that can therefore repeatedly be applied in the diagram. The concept of motif stamps aims to simplify similar actions occurring on multiple occasions in a network-drawing process.

The way the layout pattern and motif stamps can be employed in a network layout process is demonstrated in an exemplary network drawing work flow. The innovative approaches are finally rated in a user case study. The case study demonstrates that both techniques improve the drawing efficiency by accelerating the manual processes. Simultaneously, the resulting diagrams preserve recognition value and are, in contrast to automatically generated diagrams, aesthetically very pleasing.

Outlook

The presented semi-automatic layout concepts serve as a basic technology for the creation of comprehensive network diagrams. Because the approach itself is capable of visualizing all kind of networks, it would be interesting to transfer and adapt the semi-automatic layout

techniques to other application domains such as economical or social or networks.

Further work will concentrate on information visualization in the context of biochemical networks. In systems biology, simulation and experimental studies continuously produce a wide variety of data which includes but is not limited to substance concentrations, thermodynamic potentials, metabolite fluxes, reaction mechanisms, rate constants, regulatory relationships, and strengths. Due to constantly improving high-throughput techniques, we increasingly face the challenge of visualizing manifold and extensive data within large biochemical network diagrams. Large-scale exploration of time series data, and the detection and interpretation of novel properties and of system-level patterns within biochemical networks, are current issues as well as strongly compartmentalized metabolic networks. Here, a network-integrated visualization of multi-omics data sets in two and three dimensions is one of our future challenges. Furthermore, interoperability of the drawing and visualization tool Omix with databases, simulators, analysis tools, and other applications are of great importance.

Acknowledgement

We thank Andreas Kolb for constructive comments on the manuscript. Furthermore, we acknowledge all participants of the case study and the jury members as well as the anonymous referees for providing valuable suggestions.

Funding

We are grateful to Evonik Industries for financial support as part of the BMBF co-funded SysMAP project (project no. 0313704) and the EU-funded SysInBio project (project no. 212766).

References

- Gottschalk, G. *Bacterial Metabolism*. Springer-Verlag, New York, Heidelberg, Berlin, 1995.
- Michal G. *Biochemical Pathways*. An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons, New York, 1999.
- Noack S, Wahl A, Qeli E, et al. Visualizing regulatory interactions in metabolic networks. *BMC Biology* 2007; 5(1): 46.
- Rost U and Kummer U. Visualisation of biochemical network simulations with simwiz. *IEE Systems Biology* 2004; 1: 184–189.
- Qeli E. Information visualization techniques for metabolic engineering. PhD Thesis: Computer Science, University of Marburg, 2007.
- Droste P, Noack S, Nöh K, et al. Customizable visualization of multi-omics data in the context of biochemical networks. *Proceedings of the Second International Conference in Visualization*. Los Alamitos, CA, USA: IEEE Computer Society, 2009, pp.21–25.
- Neuweger H, Persicke M, Albaum S, et al. Visualizing post genomics data-sets on customized pathway maps by ProMeTra - aeration-dependent gene expression and metabolism of *Corynebacterium glutamicum* as an example. *BMC Systems Biology* 2009; 3(1): 82.
- Hu Z, Mellor J, Wu J, et al. Visant: data-integrating visual framework for biological networks and modules. *Nucleic Acids Res* 2005; 33: W352–W357.
- Stryer L. *Biochemistry*. W.H. Freeman Company, 1995.
- Suderman M and Hallett M. Tools for visually exploring biological networks. *Bioinformatics* 2007; 23(20): btm401.
- Sugiyama K, Tagawa S and Toda M. Methods for visual understanding of hierarchical systems. *IEEE Transact Syst Man Cybern* 1981; 11(2): 109–125.
- Eades PA. A heuristic for graph drawing. *Congressus Numerantium* 1984; 42: 149–160.
- Fruchterman TMJ and Reingold EM. Graph drawing by force-directed placement. *Software Pract Experience* 1991; 21(11): 1129–1164.
- Li W and Kurata H. A grid layout algorithm for automatic drawing of biochemical networks. *Bioinformatics* 2005; 21(9): 2036–2042.
- Bourqui R, Auber D, Lacroix V, et al. Metabolic network visualization using constraint planar graph drawing algorithm. *IV '06: Proceedings of the 10 International Conference on Information Visualisation (IV'06)*. London, UK: IEEE Computer Society, 2006, pp.489–496.
- Rohrschneider M, Heine C, Reichenbach A, et al. A novel grid-based visualization approach for metabolic networks with advanced focus & context view. *Proceedings of the 17th International Symposium on Graph Drawing (GD '09)*. Springer: LNCS, 2009.
- Saraiya P, North C and Duca K. Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. *Inf Visualization* 2000; 4(3): 191–205.
- Schwikowski B, Uetz P and Fields S. A network of protein-protein interactions in yeast. *Nat Biotechnol* 2000; 18(12): 1257–1261.
- Batagelj V and Mrvar A. Pajek – analysis and visualization of large networks. *Graph Drawing, Volume 2265 of Lecture Notes in Computer Science*. Heidelberg: Springer Berlin, 2002, pp.8–11.
- Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003; 13(11): 2498–2504.
- Gehlenborg N, O'Donoghue SI, Baliga NS, et al. Visualization of omics data for systems biology. *Nat Methods* 2010; 7: S56–S68.
- Villéger AC, Pettifer SR and Kell DB. Arcadia: a visualization tool for metabolic pathways. *Bioinformatics* 2010; 26(11): 1470–1471.
- Breitkreutz BJ, Stark C and Tyers M. Osprey: a network visualization system. *Genome Biol* 2003; 4(3): R22.
- Purchase HC. Which aesthetic has the greatest effect on human understanding? In: Battista GD (ed.) *Proceedings of the 5th International Symposium on Graph Drawing, Lecture Notes in Computer Science* London, UK: Springer, 1997, pp.248–261.
- Le Novere N, Hucka M, Mi H, et al. The systems biology graphical notation. *Nat Biotech* 2009; 27(8): 735–741.
- Droste P, von Lieres E, Wiechert W, et al. Customizable visualization on demand for hierarchically organized information in biochemical networks. In: Reneta PBarneva, et al (eds) *Computational Modeling of Objects Represented in Images, volume 6026 of Lecture Notes in Computer Science*. Verlag: Springer, 2010, pp.163–174.
- Hucka M, Finney A, Sauro HM, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003; 19(4): 524–531.