# On Open Problems in Biological Network Visualization

Mario Albrecht[1], Andreas Kerren[2], Karsten Klein[3], Oliver Kohlbacher[4],
Petra Mutzel[3], Wolfgang Paul[3][*], Falk Schreiber[5], and Michael Wybrow[6]

[1] Max Planck Institute for Informatics, Saarbrücken, Germany
`mario.albrecht@mpi-inf.mpg.de`
[2] School of Mathematics and Systems Engineering (MSI), Växjö University, Sweden
`andreas.kerren@vxu.se`
[3] Faculty of Computer Science, Dortmund University of Technology, Germany
`{karsten.klein, petra.mutzel, and`
`wolfgang.paul}@cs.tu-dortmund.de`
[4] Center for Bioinformatics, Eberhard Karls University Tübingen, Germany
`oliver.kohlbacher@uni-tuebingen.de`
[5] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben and Institute
for Computer Science, Martin-Luther-University Halle-Wittenberg, Germany
`schreibe@ipk-gatersleben.de`
[6] Clayton School of Information Technology, Monash University, Australia
`michael.wybrow@infotech.monash.edu.au`

**Abstract.** Much of the data generated and analyzed in the life sciences can be interpreted and represented by networks or graphs. Network analysis and visualization methods help in investigating them, and many universal as well as special-purpose tools and libraries are available for this task. However, the two fields of graph drawing and network biology are still largely disconnected. Hence, visualization of biological networks does typically not apply state-of-the-art graph drawing techniques, and graph drawing tools do not respect the drawing conventions of the life science community.

In this paper, we analyze some of the major problems arising in biological network visualization. We characterize these problems and formulate a series of open graph drawing problems. These use cases illustrate the need for efficient algorithms to present, explore, evaluate, and compare biological network data. For each use case, problems are discussed and possible solutions suggested.

## 1 Introduction

In recent years, the improvement of existing and development of novel high-throughput techniques have led to the generation of huge data sets in the life sciences. Since manual analysis of this data is costly and time-consuming, scientists are now turning towards computational methods that support data analysis.

The visualization and the visual analysis of biological networks are one of the key analysis techniques to cope with the enormous amount of data. In particular, the layout of networks should be in agreement with biological drawing conventions and should

---

[*] Corresponding author

draw the user's attention to relevant system properties that might remain hidden otherwise. While the approaches and the expertise of the bioinformatics, information visualization, and graph drawing communities may be ideally suited for solving these problems, little research has been performed to solve the special layout and visualization problems arising in the life sciences. Currently, most of the available software systems for the visual analysis of biological networks (e.g., CellDesigner [8], Cytoscape [5], VANTED [10], see also the review in [24]) provide only implementations of standard graph-drawing algorithms such as force-directed or layered approaches. Nevertheless, there are also some tools that offer specialist drawing algorithms more suitable for applications in the life sciences [2, 3, 6, 7, 13].

In general, visualization methods for the life sciences should allow for the layout and navigation of biological networks for both their static presentation as well as their interactive exploration. Such methods need to adhere to constraints that originate from recognized textbook and poster layouts (like [16, 18]), from generally accepted drawing conventions within the life-science community as well as from standardization initiatives such as MIM (Molecular Interaction Maps) [14] and SBGN (Systems Biology Graphical Notation) [19].

In this paper, we want to identify graph drawing problems originating in applied bioinformatics and network biology. We start by presenting a characterization of common biological networks, describing their structure and semantics as well as the mapping of data onto network elements. Afterwards, we present a selection of use cases describing typical uses of biological network visualization. For each use case, we present the problem as well as its relevance and discuss existing or possible straightforward solutions as well as their drawbacks. The last section will present some conclusions. The focus of this paper is less on presenting novel methods for these use cases, but more on giving an overview about open problems to raise the awareness for the manifold tasks in information visualization and graph drawing related to biological networks.

## 2 The Nature of Biological Network Data

Biological networks are used to communicate many different types of data. These data can be encoded in the structure of the network as well as represented by the network layout, or as graphical or textual annotations. The data itself may be primary data (i.e., directly measured), secondary data (i.e., derived, inferred, or predicted), or a mixture of both. In this section, we discuss some common biological networks and the types of attributes used to annotate them. Throughout the rest of this article, graphs will play an important role when trying to represent biological networks. In fact, we will use the terms *graph* and *network* synonymously when talking about the representation of biological networks.

### 2.1 Types of Biological Networks

In the following subsection we will review various types of biological networks. This listing of networks is by no means complete and includes only those networks that are most central to research in systems biology.
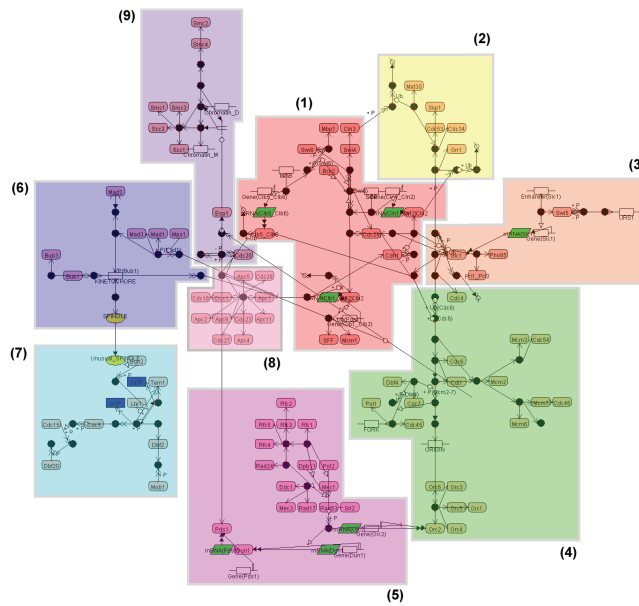
**Fig. 1:** A regulatory network representing the yeast cell cycle. The picture is taken from the CADLIVE homepage (http://www.cadlive.jp/) and was originally published in [15]. Li and Kurata used their implementation of a grid-layout algorithm.

*Gene-regulatory and signal-transduction networks* use sets of directed edges to convey a flow of information. While gene regulation (regulation of gene expression) occurs within a cell and represents a regulatory mechanism for the creation of gene products (RNA or proteins), signal transduction refers to any process that transports external or internal stimuli (often via signal cascades) to specific cellular parts where a cell response is triggered (e.g., gene regulation). While nodes in these networks represent molecular entities (genes, gene products, or other molecules), edges represent a flow of information (regulation or passing of a chemically encoded signal). Figure 1 gives an example layout of a graph representing some part of a gene regulatory network.

*Protein Interaction Networks* represent physical interactions of proteins with each other or with other binding partners such as DNA or RNA. The nodes in such networks represent proteins or sets of proteins. The time scale of protein interactions ranges from very short, transient processes (for instance, pairwise protein interactions and phosphorylation or glycosylation events) to very long lasting, permanent formation of protein assemblies (protein complexes) working as molecular machines. The interaction edges are normally undirected, but may be directed in case of specific interaction (such as activation) or of heterogeneous networks (e.g., protein-protein and protein-DNA/RNA interactions), resulting in mixed graphs. Each node and edge may be annotated with additional biological attributes like expression level, cellular localization, and the number of interaction partners. For an example layout of a protein network, see Figure 2.
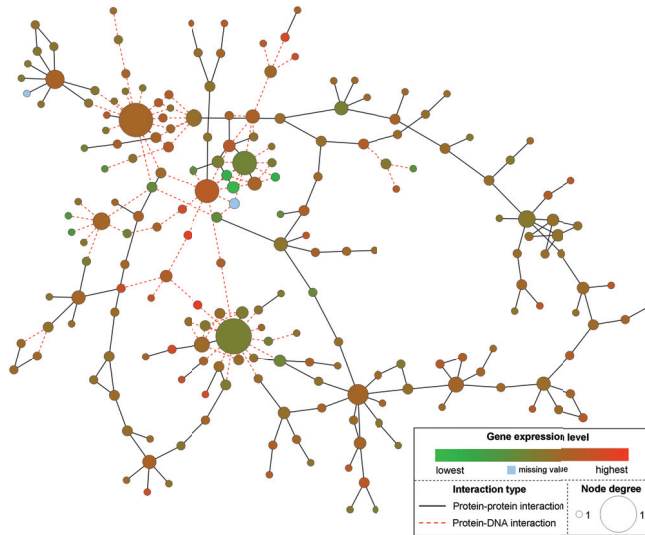
3

**Fig. 2:** Visual representation of the GAL4 protein interaction subnetwork in yeast. The protein nodes are colored by a shade gradient according to the expression value; green represents the lowest, red the highest value, and blue a missing value. The node size corresponds to the number of interactions. The shades and styles of the edges represent different interaction types; solid lines indicate protein-protein, and dashed lines protein-DNA interactions. The graph was drawn with Cytoscape [5] using its implementation of the spring-embedder algorithm.

*Metabolic Networks* describe how metabolites (chemical compounds) are converted into other metabolites. Such a network is a hypergraph that is usually represented as a bipartite graph $G = (V_1 \cup V_2, E)$. The node set is partitioned into the set $V_1$ of metabolite and enzyme nodes (enzymes catalyze the chemical reactions converting metabolites) and $V_2$ the set of reaction nodes. Large network posters (e.g., Nicholson's [18] and Michal's [16] pathway maps) are available, and several projects created graphical representations of metabolic networks and offer access to these graphs via web pages (e.g., Kyoto Encyclopedia of Genes and Genomes (KEGG) [21] or the BioCyc collection [11]). The availability of these representations has established a de facto standard for metabolic network drawings. These near-orthogonal drawings possess several characteristic features, i.e., the main direction of reaction pathways is accentuated, relevant subgraphs are placed close to the center of the drawing, substances and products of a reaction are clearly separated, and co-substances are placed out of the main path close to the reaction. Layout algorithms such as [12, 22, 23] obey established drawing styles of these networks. Figure 3 shows an example layout of a metabolic network.

## 2.2 The Attributes of Network Elements

The visual representation of primary and secondary data that has been mapped onto the elements of a biological network is an important research field. This is mainly due to the following two facts: firstly, manual analysis of primary and secondary data
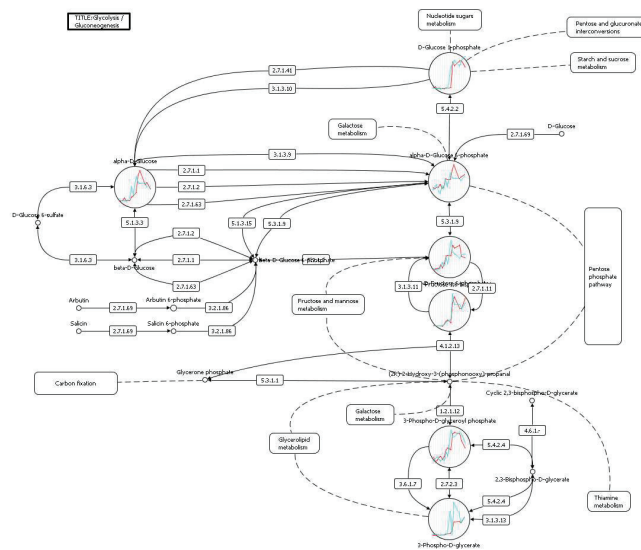
4

**Fig. 3:** Part of the glycolysis/gluconeogenesis pathway with additional data mapped onto some nodes. Circles encode metabolites, rectangles represent enzymes catalyzing the reaction, and rectangles with rounded corners denote other pathways. Solid and dashed lines represent reactions and connections to other pathways, respectively. The pathway data was derived from KEGG [21], and the graph was drawn with VANTED [10] in a style similar to the KEGG pathway picture.

has become virtually impossible as high-throughput analysis technologies have become widely available in all the life sciences. Secondly, in recent years, it has become evident that a deeper understanding of complex biological system such as cells, tissues or even whole organisms can only be achieved if the findings of life sciences such as proteomics or genomics are put into context.

The various types of primary data are defined by the different types of biochemical entities and experiments (e.g., time-series experiments or differential studies) and entities that are the subjects of analysis, namely, gene sequences, transcripts, expression levels, proteins, protein concentrations, metabolites, metabolite concentrations, or fluxes (of mass or information). The structure of the class of secondary data is, however, even more complex, and different categories of inferred, derived, or predicted information can be distinguished such as results of correlation analysis or comparisons of different biological states (e.g., healthy vs. diseased, before vs. after treatment with a drug, different organisms). The data belongs to different data types, namely:

- nominal data such as sequence names or categories
- ordinal data such as ontologies, rankings, or partly ordered information
- scalar data such as comparisons or ratios
- categorized spatial data such as data points that refer to biological entities from various parts or substructures of a cell.

The distinction between these categories is not clear in all cases. It should also be mentioned that primary as well as secondary data are subject to uncertainties (measurement errors, prediction confidences, etc.) and that the visualization of these uncertainties is often desired.

## 3  Use Cases and Related Visualization Problems

This section contains a collection of typical use cases that arise in constructing and viewing networks representing life-science data. Along with each use case, we present a formalized description of the graph-drawing, information-visualization, or visual-analytics problem behind the use case. Again, the list should not be interpreted as a closed set but rather as an attempt to stimulate further research and to demonstrate that there are many interesting and important visualization problems within the life sciences.

### 3.1  Visual Analysis of Data Correlation

Correlation graphs are frequently used as a means for visually expressing and exploring complex forms of correlation within data. As a means for visually exploring data sets within a life science context, the information contained within a data set is frequently mapped as annotation onto a graph that represents one or more types of pathways. Investigators are then interested in a graphical representation that highlights the inter-relation between the connectivity structure of pairs or subsets of nodes in the original network and their correlation. An example for an interesting correlation pattern would be a set of nodes that is closely connected within the underlying graph but exhibits only weak correlation in the data or vice versa. The represented connectivity structure should include only statistically significant correlations, for instance, significant up- or down-regulation of co-expressed genes or proteins. In particular, two or more nodes representing biological entities with multiple annotations may be considered to correlate if a minimum number of node annotations correspond with each other, e.g., regarding genotype, time value or number of the biological replicates.

One possible way to attack this problem would be to model the correlation data as a weighted graph. A weighted graph associates a label or weight with every edge in the graph. The weights are typically integers or real numbers, as this is usually a requirement induced by the algorithm applied to the weighted graph. We thus have a given graph $G_1 = (V, E_1)$ (called *network* in order to distinguish it from the correlation graph) with correlation data that induces a second graph $G_2 = (V, E_2)$ with edge weights on the same set of nodes $V$. This gives us a simultaneous embedding problem.

In our case, the two given graphs typically do not have too many edges in common. We seek either a layout of the union graph $G = (V, E_1 \cup E_2)$, in which the given network $G_1$ and the correlations are clearly displayed, or two disjoint layouts, in which the coordinates of the nodes in both layouts are identical. In the first case, a challenging task is to provide a layout which clearly emphasizes the two different edge sets $E_1$ and $E_2$. Often, a layout $\pi_1$ of the graph $G_1$ is given that has to be preserved as closely as possible. In this case, there is a trade-off between preserving the mental map (i.e., each

user's individual mental representation of a graph and his or her landmarks) and emphasizing the correlation structure. Possible solutions may either fix the layout given in $\pi_1$, or try to preserve the mental map by keeping the orthogonal relations, the topological embedding (i.e., each node of the graph is mapped to a point in the plane and each edge is associated with a curve segment between the points of its two end nodes), or the layout of a backbone (a subgraph of the given graph that serves as an abstraction).

### 3.2 Visual Comparison of Similar Biological Networks

Conservation of biochemical function during evolution results in structurally similar molecular subnetworks across different organisms and species. Uncovering relevant similarities and differences or comparing networks in different states (e.g., diseased vs. healthy), at different time points, or under various environmental conditions (for example, temperature, pressure or substrate concentrations) supports the life-science experts' knowledge-discovery process, for example, by identifying disease-specific patterns (biomarker discovery).

Given a set of graphs $G_1, \ldots, G_k$ with a high degree of similarity between each other, the task is to lay them out in a way so that the differences (or the similarities) are highlighted. This problem can be attacked via simultaneous embedding, which requires obtaining either a single layout of the union graph $G = G_1 \cup \ldots \cup G_k$ or $k$ disjoint layouts of the graphs $G_i$ $(i = 1, \ldots, k)$ such that the coordinates of all nodes common to two or more subgraphs are the same. An alternative representation has been given in [4] where the third dimension has been used to stack the $k$ layouts above each other. In these layouts, crossings between edges belonging to different graphs $G_i \neq G_j$ are either completely ignored or counted as less important than "real" crossings. In the context of biological networks and pathways, the stronger simultaneous embedding problem with fixed edges occurs, which forces not only the nodes but also the edges occurring in two or more graphs to be drawn identically. This guarantees that identical subnetworks have an identical layout. Sometimes, it may be important to keep a mental map of already given layouts of some of the graphs or their backbone structure. In any case, the layouts must obey the given biological constraints concerning the specific network type. Sometimes, the networks may be large, and it becomes desirable to hide some parts of the network and to highlight only the specific points of interest. Points of interest may be differences between the networks, but could also be important network structures such as the main pathways in a metabolic network. Here, one possibility would be to generate layouts in which the differences are all concentrated within only a few layout areas, which, of course, represents an abstraction and fails to preserve a mental map.

### 3.3 Integrated Representation of Multiple Overlapping Networks

The different types of biological networks describe different functional aspects of the whole cell, tissue, or organism in question. To get a deeper, system-wide understanding, these networks need to be combined. For example, the enzymes acting in metabolic networks are proteins and take part in protein interaction networks. It is thus becoming increasingly common to integrate these different types of networks into joint networks.

A good joint layout of these networks should reveal the *interaction* between these networks, for example, how specific nodes of the gene regulatory network activate or inactivate whole subnetworks of the metabolic network. In order to simplify the identification of these subnetworks, mental map preservation on the level of the metabolic network is helpful.

We need a representation of combined networks in which the conventional layouts (there may be several different ones) of each of these networks need to be respected. Moreover, some groups of nodes in one network may belong to groups of nodes in another network. This mapping (which may be a $1 : 1$, $1 : n$, or $n : m$ mapping) needs to be displayed in the layout. We consider the case of integrating two networks, in which the involved mapping partners can be viewed as a cluster in a cluster graph $C = (G, T)$ of cluster depth 2. Using this approach, the problem may be attacked via the following formalized problem.

We are given two cluster graphs $C_1 = (G_1, T_1)$ and $C_2 = (G_2, T_2)$ with $G_i = (V_i, E_i)$ $(i = 1, 2)$ and cluster depth 2, and a mapping function $\Phi : \mathcal{C}_1 \to \mathcal{C}_2$, where $\mathcal{C}_i$ denote the clusters in $G_i$. Generate a layout $\pi(G)$ of the union graph $G = G_1 \cup G_2 \cup G[F]$, where $F$ denotes the edges induced by the mapping, respecting the clusters as well as the conventional layouts of each of these networks. In the simpler case when the graph $G_2$ is highly disconnected (e.g., $E_2 = \emptyset$), we may prefer a solution in which the connected components of $G_2$ are integrated into a layout of $G_1$. In this case, we do not require to respect the root clusters in the problem mentioned above. Sometimes, the layout of $G_1$ may be given. If this is the case, there is often a trade-off between sticking to the given layout as closely as possible (or trying to preserve a user's mental map or the backbone of the original graph) and obtaining a better representation of $G_2$.

### 3.4 Visualization of Sub-cellular Localization

Cells consist of distinct compartments, subcellular locations, separated from each other by membranes. Examples for these are the cytosol, the nucleus, the mitochondria, or chloroplasts in plants. The membranes enclosing a compartment separate parts of the biological networks as well. Different partitions of the network will be localized in different subcellular locations and hence cannot interact with each other directly. It is thus essential for an understanding of the network's function to integrate that spatial information into the layout of the network. The required localization data may be either already contained in data sets derived from experiments, may be extracted from external sources, or may be predicted.

Given a network $G = (V, E)$ and additional localization annotation for the nodes in $V$, we search for a layout that reflects the topographical information of $G$ and that conforms to the drawing conventions for that type of network. It should, of course, be at the same time aesthetically pleasing. Note that the subcellular localization does not just give a clustering of the nodes, for example, a specific relative position of the cellular compartments may be implicitly given by the biological morphology. If $G$ reflects a flow of mass or information, the direction of the flow also needs to be displayed by, e.g., hierarchical layering of the nodes. A simple representation of a cross-sectional cut through a cell would be a stacking of layers as in [2]. In a layout that is inspired by biological morphology, subgraphs of the network should be arranged according to their

subcellular localization in such a fashion that the physical structure of the compartments delimiting these subgraphs becomes evident [6, 17]. In order to increase the user's acceptance of such layouts, it may be necessary to resemble the (manually generated) layouts from established projects like BioCarta [1]. Using a layered drawing approach has the inherent flaw that mass and information may only pass freely between members of neighboring layers.

### 3.5  Visualization of Multiple Attributes

Analyzing data in a life-science context often requires the consideration of multiple data attributes. One example for this would be the analysis of several time-series data sets in the context of a biological network. This is frequently done in order to better understand the dynamic behavior of a biological system. The combined representation of such time-series data and a corresponding network should allow investigators to gain new insights concerning the underlying system, such as co-regulated elements and their connection within the network. Such an analysis can, for example, be achieved by mapping the data onto the nodes of a network, see Figure 3 where time-series data from two series (day and night) were mapped on parts of the network (cp. enlarged nodes).

Mapping the given data onto nodes and/or edges is one possibility of solving the problem. It is also the simplest and most straight-forward one but has the general problem that simple use of small visualizations that replace, for example, the node representations is most of the time not sufficient, as a visual comparison of such small graphics in a large graph becomes very quickly infeasible. Furthermore, one has to consider that the attributes involved may belong to various types of data. Each data type needs specific requirements for its visual representation [9]. The challenge here lies in finding a harmonic combination of visual representations for mixtures of different data types.

A possible solution lies in the use of so-called preattentive features [9]. A preattentive feature is a visual property of a picture or drawing that can be very rapidly and accurately detected by the low-level human visual system. Therefore, users do not have to focus their attention onto one specific part of a drawing in order to understand its basic visual properties. Another possibility is to use an additional view separated from the original network view for displaying the attributes. This idea is based on standard coordinated and multiple view visualization techniques. Drawbacks of this method are the need to add connections between the views as well as the spatial distance between both. This is usually done using brushing-and-linking techniques that are used to highlight, select or delete subsets of elements by pointing to specific elements. When multiple views of the same data are used simultaneously, brushing is typically associated with linking, i.e., brushing of elements in one view affects the same elements in other views. First results that try to compare both approaches [25], i.e., multiple views and attribute integration, cannot be directly used in our complex case of biochemical networks.

### 3.6  Visualization of Flows and Paths in Networks

The qualitative and quantitative distribution of mass and signal flows (fluxes) within a biological network has to be analyzed under consideration of uncertainties in the data. The flow along certain paths may change over time (time-series of measurements) and

the paths through the network may be numerous so that not all of them can be displayed. Investigators are, in such cases, primarily interested in the main paths through the network, i.e., those paths that possess a statistically significant flow and transport a considerable percentage of the overall flow through the entire network, and in the metabolites and reactions that are involved in these paths.

The given network, together with quantitative and qualitative information about the flow of mass or information (edge weights), may be a potentially very large one. For directed graphs such as metabolic networks, the layout must reflect the hierarchical nature of the flow, preserve layouts for subnetworks originating from textbook representations as closely as possibly, adhere to general drawing conventions, and, at the same time, focus on the relevant parts of the network, e.g., paths that at a certain point in time transport a large part of the flow. These main paths thus have to be visually emphasized (e.g., placed at the center of the layout and drawn as straight lines) and the distribution of the fluxes within the network have to be depicted, for example, by using different edge widths or colors.

If the dynamic change in the flow over time also needs to be visualized, smooth animations between layouts are required to preserve the user's mental map. In the past, this problem was mainly covered by software visualization techniques, especially by algorithm animation. In information visualization, solutions for showing flows and paths in networks are relatively rare and mostly limited to special cases or domains, to visualize communication flows in social networks [20].

### 3.7 Exploration of Hierarchical Networks

Biological networks often comprise several thousands nodes and edges. To support the exploration of such large and complex structures, the entire network is usually broken down in a hierarchical manner into pathways and subpathways. Investigators commonly focus on (sub)pathways in a region of interest and explore their relation to other pathways. However, due to the many connections between different pathways, an abstract overview-like picture of all pathways interconnections as well as an interactive navigation from a set of pathways to other connected or related pathways is often desired.

Given a huge biological network, methods for the biologically meaningful visualization of selected subsets $G_1, \ldots, G_k$ (e.g., pathways) and their interrelations, as well as techniques for the navigation within the network are needed. In order to allow the user to keep his orientation during exploration of the network, the layout changes resulting from a user interaction (e.g., selection of an additional pathway) should be small, and context information needs to be represented in an appropriate way. Expand-and-collapse mechanisms thus need to be incorporated into layout algorithms such that drawing conventions and the mental map are preserved. These operations could be restricted to certain levels of abstraction, e.g., by only collapsing/expanding semantically meaningful substructures like pathways. One of the main challenges is that layouts for such subnetworks as well as their relative position to each other may be pre-specified. This layout information needs to be preserved as closely as possible. As these networks are too large to be laid out nicely as a whole, an overview graph or a representation of the backbone could be defined by reduction or abstraction that covers the topologically

or semantically relevant features of the network, thus helping the user in navigating through the network.

The subsets $G_1, \ldots, G_k$ do not need to be disjoint but may partially overlap. This poses an additional challenge for the visualization problem: Either the duplicate components are merged, which complicates the task of mental map preservation, or it has to be clearly emphasized that they represent the same biological entity.

## 4   Conclusions

Biological networks play a crucial role in systems biology. Many universal as well as special-purpose tools and libraries are available for laying out and drawing graphs in order to help visually investigating these networks. However, these tools either do not adhere to the special drawing conventions and recognized layouts in biology or are not adequate for handling large graphs.

The use cases presented here reveal graph drawing as well as information visualization problems arising in the biological domain. While we present possible solutions to these problems, we also consider this paper a challenge to the graph drawing community and people working on network visualization in systems biology as well as the life sciences. The problems described here are far from being solved for all practical scenarios and certainly merit further attention. Developing improved solutions will require custom state-of-the-art graph-drawing approaches, and more importantly, collaboration between researchers from graph drawing, information visualization, visual analytics, and the life sciences. We hope that this paper encourages such collaborations and presents interesting research directions in these fields.

## Acknowledgements

## References

1. BioCarta. http://biocarta.com/.
2. A. Barsky, J. Gardy, R. Hancock, and T. Munzner. Cerebral: a cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. *Bioinformatics*, 23(8):1040–1042, 2007.
3. M. Becker and I. Rojas. A graph layout algorithm for drawing metabolic pathways. *Bioinformatics*, 17(5):461–467, 2001.
4. U. Brandes, T. Dwyer, and F. Schreiber. Visualizing related metabolic pathways in two and a half dimensions. In *Graph Drawing*, volume 2912 of *LNCS*, pages 111–122, 2003.

5. M. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, and K. Hanspers et al. Integration of biological networks and gene expression data using Cytoscape. *Nature Protocols*, 2(10):2366–2382, 2007.

6. E. Demir, O. Babur, U. Dogrusöz, A. Gürsoy, G. Nisanci, R. Çetin Atalay, and M. Ozturk. PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics*, 18(7):996–1003, 2002.

7. U. Dogrusöz, E. Giral, A. Cetintas, A. Civril, and E. Demir. A compound graph layout algorithm for biological pathways. In *Proc. International Symposium on Graph Drawing (GD'04)*, LNCS, pages 442–447. Springer Berlin, 2004.

8. A. Funahashi, M. Morohashi, and H. Kitano. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico*, 1(5):159–162, 2003.

9. C. Görg, M. Pohl, E.Qeli, and K. Xu. Visual Representations. In *Human-Centered Visualization Environments*, volume 4417 of *LNCS*, pages 163–230. Springer, 2006.

10. B. 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.

11. P. Karp, C. Ouzounis, C. Moore-Kochlacs, L. Goldovsky, P. Kaipa, D. Ahren, S. Tsoka, N. Darzentas, V. Kunin, and N. Lopez-Bigas. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Research*, 33:6083–6089, 2005.

12. P. Karp and S. Paley. Automated drawing of metabolic pathways. In *Proc. International Conference on Bioinformatics and Genome Research*, pages 225–238, 1994.

13. P. Karp, S. Paley, and P. Romero. The pathway tools software. *Bioinformatics*, 18(S1):S225–S232, 2002.

14. K. Kohn and M. Aladjem. Circuit diagrams for biological networks. *Molecular Systems Biology*, 2:e2006.0002, 2006.

15. W. Li and H. Kurata. A grid layout algorithm for automatic drawing of biochemical networks. *Bioinformatics*, 21(9):2036–2042, 2005.

16. G. Michal. *Biochemical Pathways, 4th edition (Poster)*. Roche, 2005.

17. M. Nagasaki, A. Doi, H. Matsuno, and S. Miyano. Genomic Object Net: a platform for modeling and simulating biopathways. *Applied Bioinformatics*, 2:181–184, 2004.

18. D. Nicholson. *Metabolic Pathways Map (Poster)*. Sigma Chemical Co., 1997.

19. N. L. Novère, M. Hucka, H. Mi, S. Moodie, F. Schreiber, A. Sorokin, E. Demir, K. Wegner, M. Aladjem, S. Wimalaratne, and F.T. Bergman et al. The Systems Biology Graphical Notation. *Nature Biotechnology*, 27(8):735–741, 2009.

20. D. Offenhuber and J. Donath. Comment Flow: visualizing communication along network paths. Poster presented at IEEE InfoVis '07, 2007.

21. H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 27:29–34, 1999.

22. F. Schreiber. High quality visualization of biochemical pathways in BioPath. *In Silico Biology*, 2(2):59–73, 2002.

23. M. Sirava, T. Schäfer, M. Eiglsperger, M. Kaufmann, O. Kohlbacher, E. Bornberg-Bauer, and H. Lenhof. BioMiner - modeling, analyzing, and visualizing biochemical pathways and networks. *Bioinformatics*, 18(S2):219–230, 2002.

24. M. Suderman and M. Hallett. Tools for visually exploring biological networks. *Bioinformatics*, 23(20):2651–2659, 2007.

25. B. Yost and C. North. Single complex glyphs versus multiple simple glyphs. In *CHI '05 extended abstracts on human factors in computing systems*, pages 1889–1892. ACM, 2005.