DOI: 10.1111/cgf.14828
Eurographics Conference on Visualization (EuroVis) 2023
D. Archambault, R. Bujack, and T. Schreck
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 42 (2023), Number 3

visMOP – A Visual Analytics Approach for Multi-omics Pathways

N. Brich[1], N. Schacherer[1], M. Hoene[2], C. Weigert[2], R. Lehmann[2], M. Krone[1,3]

[1]Institute for Bioinformatics and Medical Informatics, University of Tübingen, Germany
[2]Institute for Clinical Chemistry and Pathobiochemistry, University Hospital Tübingen, Germany
[3]Tandon School of Engineering, New York University, USA

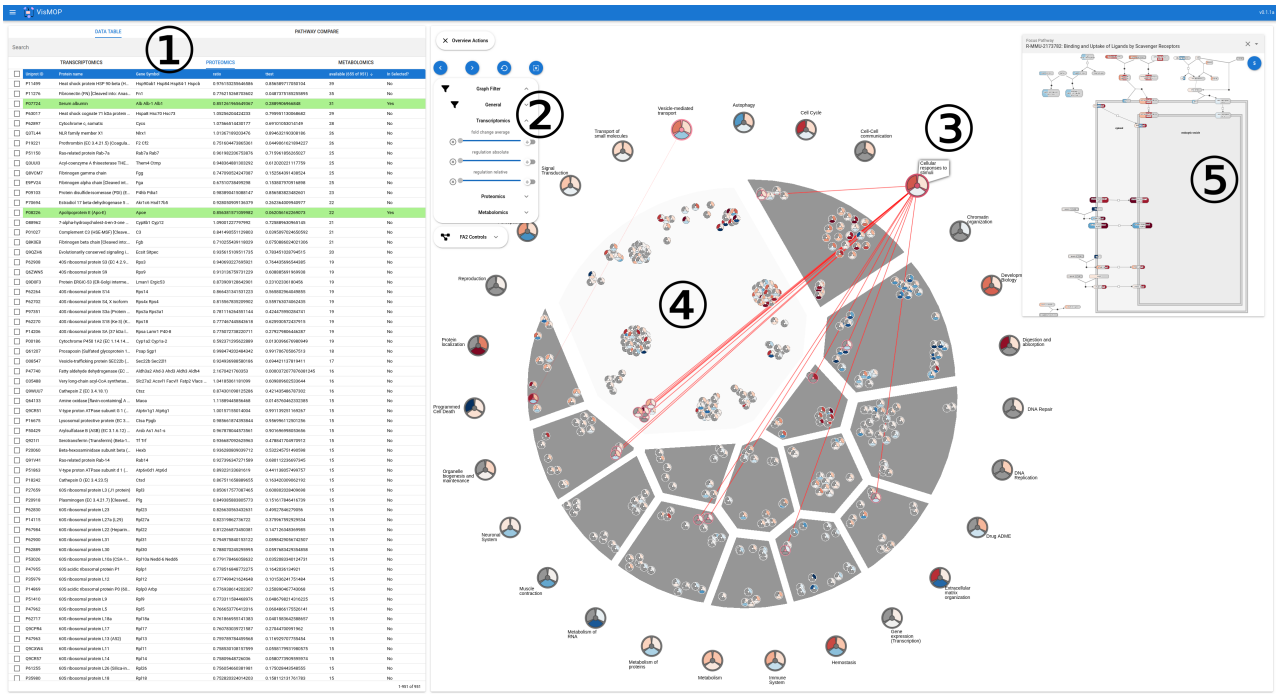


Figure 1: *The main view of visMOP (visualization of Multi-Omics Pathways) showing the overview of the cluster graph with the Reactome superpathways laid out as a circle around it (③). The dark gray regions show clusters of pathways containing similar data, while the light gray region shows pathways not clustered with other pathways (④). A selected pathway is highlighted, showing a tool-tip and both hierarchical as well as non hierarchical connections to other pathways. On the right, the detail view window is shown for a selected pathway (⑤). On the left, the visual filter options, as well as option to control the intra-cluster layouting are displayed (②). The input data is shown in a collapislbe table linked to the main overview visualization (①).*

Abstract

We present an approach for the visual analysis of multi-omics data obtained using high-throughput methods. The term "omics" denotes measurements of different types of biologically relevant molecules like the products of gene transcription (transcriptomics) or the abundance of proteins (proteomics). Current popular visualization approaches often only support analyzing each of these omics separately. This, however, disregards the interconnectedness of different biologically relevant molecules and processes. Consequently, it describes the actual events in the organism suboptimally or only partially. Our visual analytics approach for multi-omics data provides a comprehensive overview and details-on-demand by integrating the different omics types in multiple linked views. To give an overview, we map the measurements to known biological pathways and use a combination of a clustered network visualization, glyphs, and interactive filtering. To ensure the effectiveness and utility of our approach, we designed it in close collaboration with domain experts and assessed it using an exemplary workflow with real-world transcriptomics, proteomics, and lipidomics measurements from mice.

CCS Concepts

• ***Human-centered computing*** → *Graph drawings; Visualization techniques;* • ***Applied computing*** → *Bioinformatics;*

1. Introduction

Biomedical research often targets questions that require a broad perspective: for example, a single mutation inside a gene may cause a complex disease while on the other hand, apparently simple diseases are frequently caused by a multitude of genetic and non-genetic factors. For many years, such rather comprehensive research questions had to be broken down into quasi-atomic chunks to be answered with the available techniques. The last decades, in contrast, have seen a continuous increase in techniques grouped under the umbrella term *omics*, where transcriptomics is, for example, the analysis of the complete set of transcribed genes at a certain time and condition. The next step towards a comprehensive study of biological systems, which are characterized by a high degree of interconnectedness, is to combine two or more of these omics techniques.

For instance, enzymes, which are crucial for the generation of many essential metabolites, are produced during gene expression using mRNA transcripts that are, in turn, products of genes, i.e., DNA transcription. How much of a given gene is transcribed from DNA into RNA (as measured by transcriptomics), however, is not only regulated by transcription factors but by a multitude of proteins that either promote or inhibit translation [CMH*99; Lat13]. The mRNA produced in this step is further regulated by proteins that affect, e.g. its stability, and the same is true for translation, the production of proteins from mRNA. And of course, translation—the reading of the processed mRNA and the production of proteins—is also facilitated by various protein and other metabolic processes [Her89]. That is, the presence and abundance of the proteins (e.g., measured by proteomics) is an important factor. Even the metabolites themselves have been shown to play a significant role during many steps of the gene expression process and the regulation of enzyme activity [Lad06; WMWH15], therefore, metabolomics measurements also needs to be considered. Therefore, conclusions about biological function and regulation can, ultimately, only be drawn by investigating several of these layers in a combined fashion. Performing a combined, comprehensive analysis that considers the interconnectedness increases the effectiveness of and confidence in the analysis, as well as enhancing the acquired information of a given process or system [WZR*19; CSB*20]. Utilizing multiple data sources, however, leads to a growing amount of data—especially since the data volume is constantly increasing due to higher resolution, be it spatial, temporal, or depth of sampling. Simultaneously, the complexity of the data increases as well, as different methods often do not yield data that can be mapped directly to a common denominator.

While several advances in the past, like the increase in computational power and the development of improved visualization tools, facilitate the analysis of these complex data sets, working on this kind of data still remains a challenge [RS18]. Depending on the type and specificity of the research question, different approaches to interpret this data can be taken. For very specific research questions, purely algorithmic approaches might yield sufficient answers. If a more exploratory analysis is needed—for example, to formulate, refine, or reject a hypothesis based on the available data—a visual analytics approach might be more appropriate.

To enable comprehensive visual analysis, the measured omics data needs to be embedded into the relevant biological pathways (i.e., prior knowledge about reactions and interactions between molecules). Consequently, metabolic pathway visualizations that are already familiar to the domain experts are a suitable starting point for designing an advanced, effective visualization. Many of the most common metabolic pathway visualizations are, however, manually curated diagrams, which limits the flexibility to reuse this information and makes it difficult to incorporate them into a visual analytics application. On the other hand, common generic network visualization methods, like force-directed node-link diagrams, often disregard the biological background knowledge and can result in unfamiliar layouts that are hard to understand for the users.

We present a visual analytics approach that allows for exploratory analysis of multi-omics datasets. To facilitate the analysis, we propose to use a clustered graph to visualize multi-omics datasets on a pathway level while representing the pathways as glyphs encoding the contained data. Our contributions can be summarized as follows. We present an approach that works with a variable number of different omics measurements, which are mapped against all known metabolic pathways from the Reactome [GJS*22] database. The resulting pathways containing the measurement data are summarized in an overview graph, which is visualized as a node-link diagram showing the interconnections between the pathways. For this network of pathways, we present a data-driven layout algorithm that can be customized by the users according to their specific needs, interests, and research questions. The pathways, which are the nodes of the overview graph, are represented by glyphs showing the omics measurements mapped to those pathways. That is, our approach allows users to compare different pathways with respect to the measured data and facilitates the detection of correlations or mismatches in the data that require further investigation. We developed visMOP (visualization of Multi-Omics Pathways), a web-based prototype implementing our approach that also provides level-of-detail functionality for the overview graph and multiple linked views for a more detailed inspection of individual pathways. We tested our approach on a real-life dataset together with domain experts.

2. Related Work and Biological Background

This section provides links to surveys about network visualization (Section 2.1) and introduces applications of networks in biology and data sources for pathway data (Section 2.2). We also briefly discuss related visualization applications for multi-omics data (Section 2.3).

2.1. Network Visualization

The visualization of networks can help to analyze the encoded relationships and elucidating underlying structures. It has been the target of research for a considerable time and has, thus, grown into an extensive field of research. Giving a comprehensive overview of network visualization is beyond the scope of this paper. Therefore, we only give a very brief introduction and refer to state-of-the-art reports about the visualization of large [vLKS*11], dynamic [BBDW17], multivariate [NMSL19], and multilayer [MGM*19] graphs, which summarize foundational methods and recent advances in the field of network visualization and provide pointers for further reading. Koutrouli et al. [KKPP20] recently gave an overview of commonly used graph visualizations for biological networks.

There are various ways networks can be displayed, depending on

the use case. The most common representations are adjacency matrices and node-link diagrams [KEC06]. While adjacency matrices make it easy to look up direct connections between nodes and—depending on the sorting—identify cliques and clusters, they are often considered to be less intuitive, and it is harder to find paths between nodes compared to a node-link diagram. Node-link diagrams, on the other hand, highly depend on the layout, i.e., the spatial positioning of the nodes. A recent study [ASA*22] investigated and compared the performance of adjacency matrices, bipartite layouts, and node-link diagrams for five network overview tasks.

Especially for large networks with many nodes and edges, a good layout algorithm is crucial to enable users to detect important features like clusters, symmetries, and interconnected areas [KKPP20]. Many approaches to generate layouts have been proposed. The most simple layouts place the nodes on grids or circles. Many layout algorithms are based on simple simulations, e.g., modeling networks as a set of bodies connected by spring-like forces. While these approaches can be useful for analyzing smaller networks, they are often not useful for analyzing larger, more complex networks because they usually do not consider any property of the network.

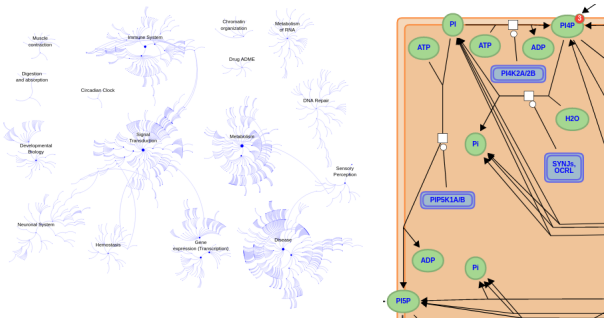


Figure 2: *Metabolic pathways visualized by Reactome [GJS*22]. Left: excerpt of the overview of the metabolic pathway network where each metabolic region is the root for a hierarchy visualized using a radial layout. Right: excerpt of a detail reaction pathway containing molecule-level information. These pathways are located in nodes closer to the leaves of the radial hierarchy.*

2.2. Biological Networks

Many disciplines in biology use networks and their visualization, most commonly node-link diagrams, to model and analyze relations between different entities e.g., phylogenetic trees showing the evolutionary relatedness of species [PDA10; HMB*18], or graphs modeling the connections between neurons in neurobiology [Pes14; YSD*17]. Molecular biology has a plethora of different application cases for networks, e.g., gene-regulatory networks [KS08; DGC*17], protein-protein-interaction graphs [NYP12; SM03], or gene-variation graphs [ENS*20; AHN*21]. The modeling of metabolic pathways is an important application area of networks in biochemistry [NDG*17; KBC*19; GJS*22]. These metabolic pathways describe how cells and, in turn, organisms function on a molecular level. They are, thus, instrumental in understanding any biological process. The nodes of the graph usually represent reactants that are connected by edges indicating reactions. Reactants can be a wide variety of biologically relevant molecules, e.g., proteins, small molecule metabolites, or DNA/RNA. As many organisms feature tens of thousands of biologically active entities (genes, proteins, RNA), and even more different types of interactions, the resulting networks are very large and complex. This makes it challenging to generate clear, meaningful visualizations of these networks. Thus, in the applied biomedical domain, many of the most well-known and widely used metabolic pathway visualizations are manually curated and laid out diagrams. Recently, a first approach for the automatic generation of metabolic pathway maps that can retain many of the desirable features of the manually laid out diagrams has been presented [WNSV19]. However, it just considers information about the known interactions and does not incorporate omics measurements, making it unfeasible for our intended use case.

Reactome [GJS*22] is a pathway database offering information about molecular entities in conjunction with metabolic reaction pathways. The pathways are arranged in a hierarchy: each pathway is part of a multilevel hierarchy in which general metabolic topics constitute the roots and fine-grained reaction sequences constitute the leaves. The more complex pathway diagrams found at the roots and the levels further removed from the leaves are created manually (Fig. 2). Reactome is part of the *ELIXIR* [CT12] initiative and offers its data and software under a creative commons license. This includes the instructions of all pathway diagram drawings, allowing a fast and efficient way of querying large amounts of data.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) [KG00] is another widely used database offering information about different molecular entities and metabolic pathways. It presents the pathways in a flat hierarchy in which they are linked by functional connections. A separate class hierarchy groups the pathways. Similar to Reactome, the pathway diagrams are created manually to consider previous knowledge and conventions for a given process the diagram describes. However, it introduced a subscription model in 2011 due to funding issues and detailed information about how to draw the pathway diagrams is not available, making it less convenient to use.

Especially among practitioners, the two *"Biochemical Pathways"* posters by Roche [Mic17] are popular. They show a comprehensive summary of known pathways and processes. The manually created graphs are structured into labeled sections on different high-level topics, e.g., *Citrate and Glyoxalate Cycle* or *Steroid Metabolism*. The lower level shows the actual chemical reactions with all participating molecules. Roche offers print and zoomable online versions of the posters, however, the data is proprietary and not publicly available, severely limiting the use for visual analytics applications.

2.3. Multi-omics Pathway Analysis Tools

In contrast to KEGG and the Roche biochemical pathway map, Reactome [GJS*22] offers an interactive visual analysis tool, which allows the user to map multi-omics measurement data onto the Reactome pathways. The visualization is rather simple: all omics measurements are aggregated and the pathway nodes, which are arranged in a hierarchical node-link diagram, are colored by either coverage or significance in coverage, or (when measurement data is supplied) the average over all supplied omics. It does not influence

the graph layout. Additionally, the data can be shown as "*Reactfoam*", a hierarchical Voronoi treemap used to show the data in lieu of the node-link diagram. Any further analysis then has to be performed on a detail pathway level.

Paintomics [LSP*22] is a tool tailored to multi-omics. It can map different omics types—including, but not limited to: transcriptomics, metabolomics, proteomics, and different regulatory and region-based omics—to pathway maps from KEGG, Reactome, and MapMan [SPK*19]. Using these data, significantly enriched pathways are searched in these databases. The user can then select the pathways, and the molecular entities in the pathway visualization are then colored according to the corresponding data value. That is, the provided overview network visualization is mainly an entry point for the analysis of individual pathways. Our approach, in contrast, aims at generating a rich overview using layout and clustering techniques to generate insights from this overview by combining multiple omics, e.g., allowing users to identify groups of similarly regulated pathways and mismatches between the different omics levels. Pathview Web [LPB*17] follows a similar approach: it simply maps colors showing the regulation to the KEGG pathway drawings.

OmicsNET [ZX18] leverages the third dimension to accommodate for the complexity of multi-omics networks. Besides the omics data, it also allows the user to load additional information like SNPs for genomics or taxon information for studies involving multiple organisms. 3D visualizations, however, have not been shown to yield additional insights without the use of stereo-vision, i.e., VR or AR [MGM*19]. Such systems are typically not widely available in molecular biological laboratories, which are our target audience. Additionally, 3D network visualizations suffer from occlusion, perspective distortion, and usually require more user interaction [Mun15]. To avoid this, we settled for 2D visualization. Furthermore, instead of focusing on details brought in by additional data, we want to show large-scale, pathway-overarching effects on the metabolism.

There are also frameworks targeted at systems biological network visualizations like VANTED [JKS06; RJH*12], which can accommodate multiple omic types. While VANTED can also provide overviews, its focus lies on analyzing and visualizing biological data on a detail level in the context of systems biology, tracking changes in biological reactions and single molecule types like metabolites.

3. Tasks and Requirements

Our domain experts currently use different tools, such as g:Profiler [RKK*19], to either perform pathway enrichment analyzes for each type of omics data individually or for subsets of features that overlap in the individual statistical analyzes of the different omics datasets. Similarly, MetaboAnalyst [PCZ*21], which allows for an integrated analysis of metabolomics and either transcriptomics or proteomics data, combines the data without allowing for an exploration of the contribution of individual features from the different omics sets. This obscures the structure of the data, which is particularly problematic given that both parallel and antiparallel responses of different omics types can be relevant, e.g., an enzyme in a pathway can be upregulated, causing a higher amount of its product, but a higher amount of its product could also decrease the amount of the enzyme in a regulatory feedback mechanism.

Thus, our explicit goal was to develop an approach to unify the analysis and visualization of multiple omic types. Our domain experts explained that they are in need of a approach to give an exploratory overview of the data at hand, which can provide an entrance point for hypothesis generation and further detail investigations. When formulating the tasks together with the experts, we considered several tasks taxonomies as a reference [MMF17; BM13]. A concrete overview task is to identify high level groupings of metabolic pathways to uncover sets of co-regulated or connected pathways. Detail investigations require an additional view on the more granular aspects of the data, requiring the framework to support various additional by offering detail-on-demand functionality. They, for example, need more details about a specific pathway on a molecular level, thus requiring a detailed pathway view, as found in the Reactome or KEGG pathway visualization. Additionally, users might want to compare different pathways to assess if similar omic patterns appear within them.

We propose a combination of a layout algorithm tailored to multi-omics data and glyphs that show information about the pathways. We combine these with a level-of-detail approach to further increase the scalability and to declutter the visualization. Both the layout and the glyph design were created with several requirements in mind. The requirement analysis was done together with four domain experts working in biochemical and biomedical research. In a series of semi-structured interviews, we first collected information about the data, analysis challenges, and common tools and workflows. From this information, we inferred an initial set of requirements, which we then discussed with the experts to ensure their validity. During this phase, we also showed several iterations of an earlier prototype [KDC*20] to the experts and reiterated on the requirements. Finally, we converged on the following set of five requirements:

R1 The visualization must be able to integrate multiple omics types and assess differences in the measured data.
R2 The approach must be able to scale up/down when more/less omics types are available.
R3 The representation must give equal importance to all the included omics types.
R4 The visualization needs to provide a summary of the individual pathways that allows for a quick comparison while also providing the opportunity for detailed inspection.
R5 The network layout should make use of established pathway layouts.

4. Visualization of Multi-omics Pathway Networks

Some of the mentioned visualizations, like the Roche or KEGG pathways, helped to establish mental maps for many metabolic processes. These visualizations, however, present information on a detailed level and are not suitable as an overview. In an earlier iteration of visMOP, we attempted to generate a global network incorporating all detail pathways utilizing the established detail layouts of KEGG, resulting in a very large graph with thousands of nodes and tens of thousands of edges. Based on the domain expert feedback, however, we realized that this approach does not yield much useful insight, which prompted us to change the focus of the application onto giving an overview of the data, with details-on-demand for individual pathways.

This left us with the challenge of uniting existing mental maps with a data-driven overview. The overview maps of KEGG or Roche, however, are not particularly well suited for overview tasks as they are high-resolution abstractions of the detail pathways, difficult to use on regular-sized screens. Reactome's more abstract overview, on the other hand, is based on a hierarchy, making it more suitable. However, it can not be easily adapted to a data-driven layout. We, thus, chose to implement a novel layout based on the Reactome hierarchy, which maintains the mental map of users familiar with the Reactome hierarchy. Additionally, we maintain the detail pathways' layout of Reactome. Consequently, users who perform a more detail analysis using visMOP can build on their previous knowledge.

Our approach combines multiple views and uses a glyph metaphor to encode pathways, as recommended by Cruz et al. [CAM19]. We use a data-driven layout approach to tailor the visualization to the underlying data. Users can interactively adjust the input data, the parameters for the data-driven layout, and the appearance of the visualization to facilitate the analysis.

4.1. Pathway Network

The multi-omics pathway overview graph (see Fig. 1, ③) consists of nodes representing different metabolic pathways obtained from the Reactome database. Relationships between the nodes—i.e., edges—correspond to two categories: They are either hierarchical, as defined by Reactome or non-hierarchical, logical connections directly inferred from known molecular interactions (e.g., "pathway A produces metabolite X which is then used in pathway B"). There are two different subtypes of pathways: *superpathways*, which represent broad metabolic topics (e.g. Signal-Transduction or Innate Immunity) and are positioned at the root of the pathway hierarchy, and non-root "regular" *pathways*, which include molecular-level information and can be found either as leaves or closer towards the leaves of the hierarchy. To create a clear visualization, we apply two different layout algorithms to these two subsets of nodes (***R5***/***R1***). For the superpathway hierarchy level, we use a simple circular layout, since there are only a few superpathways. To make it easy to find a certain superpathway, the nodes are sorted clockwise in alphabetical order by their names. The labels showing these names can be placed outwards to ensure good readability (see Fig. 1, ③). In the center of this circle of superpathway nodes, the nodes that are lower in the hierarchy are shown (see Fig. 1, ④). This is the main pathway network, consisting of pathways that include the actual molecular interactions. Here, a more complex, data-driven layout is applied, which is described in Section 4.2. It should be noted that not the complete hierarchy provided by Reactome is displayed. We limit our overview network to the first pathways (in root-to-leaf order) for which Reactome associates a molecule-level diagram (thus omitting intermediate levels between the superpathways and the actual molecular interactions).

As the network is densely connected, especially to the superpathway nodes on the surrounding circle, the edges would cause intense visual clutter if all of them would be displayed. This would make the identification of source and target nodes and paths between them very difficult. To alleviate this, edges are hidden by default. Only when a node—superpathway or pathway—is hovered, the edges connected to this node are shown.

4.2. A Data-Driven Layout Algorithm for Pathway Networks

Many node-link visualizations utilize simple layout algorithms that only consider the graph topology to generate a somewhat understandable visualization. These layouts usually minimize overlaps and reduce edge crossings; however, since they usually do not consider non-topological information, i.e., the underlying data values, they rarely result in layouts that are meaningful for the interpretation and analysis of the data, in the sense that the semantics of the data are taken into account. Therefore, we devised a layout algorithm that takes the characteristics of the underlying data into account.

Our proposed layout is based on the idea of using clusters, similar to the Grid- and Modularity-based Layout algorithm by He et al. [HLY*19]. For the clustering, the values and statistics about the supplied multi-omics data are taken into account, such as the average *fold change* per omics, the number of regulated entities per pathway, etc. (***R1***). The fold change is commonly used in bioinformatics to express the change of a given measurement value from an experimental condition—most often the control—to another one. Note that either all available parameters or only a user-defined subset can be used as input. While these metrics, available in the presented prototype, are not exhaustive, and include others (Supplement, Table 1), they serve to highlight the user customizable approach of selecting data metrics which are relevant to the analysis case. For further, more specific applications additional metrics tailored to the use case can be implemented straightforwardly.

Density-based clustering in high-dimensional Euclidean spaces can become problematic due to the sparseness of the space—a phenomenon often dubbed the "curse of dimensionality" [Ass12; HK99]. One way to overcome this is to use dimensionality reduction methods like UMAP [MHM20] as a preprocessing step [Ass12]. As an input to UMAP we use the parameters derived from the input data, which are normalized by transforming them to a Z-score. We apply UMAP twice: once, if more than two features are chosen, to generate a lower-dimensional space (with a default $N/2$-target dimensions) for clustering with OPTICS [ABKS99]. For visualization purposes time we project the data a second time, this time to two dimensions, to determine cluster centers as the basis for a next step. We chose OPTICS, because it is an unsupervised method that determines the ideal number of clusters and can classify entities as noise if they do not fit into a cluster. We run the clustering several times with different values for the minimum size of a cluster to find an ideal clustering by means of the silhouette value [Rou87] for differently-sized datasets. That is, the user does not have to parameterize the OPTICS clustering (for epsilon, we use the default value of ∞). We assign the noise entities to a virtual cluster that is marked in the final visualization to illustrate that these entities are not similar (region with light gray background; see, e.g., Fig. 1, ④).

In these next steps, the space formed by the above-mentioned high-level hierarchy nodes is partitioned in a way that both the distance from one cluster to another and the proportion of nodes contained vs. space allocated are considered. He et al. [HLY*19] use rectangular regions for the clusters in this step and try to optimize their placement to reduce edge crossings. In contrast, we use a circular layout, in which a Voronoi treemap is used to allocate space [BD05; NB12]. This results in a layout in which similar clusters are placed next to each other (see Fig. 3).

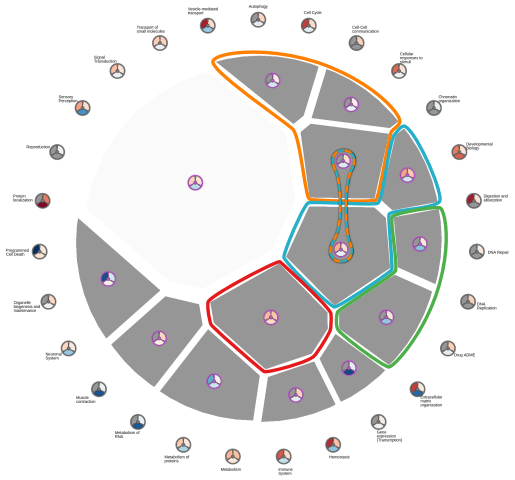
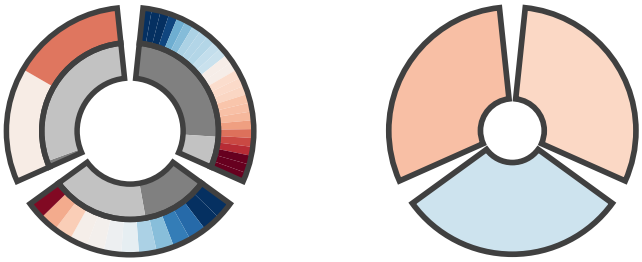


Figure 3: *visMOP overview annotated to show similar cluster aggregate nodes in groups. As the aggregated nodes indicate the average regulation of the supplied omics data of the pathways contained in the cluster, this grouping thus also groups by average regulation. The cluster regions similar to each other are placed in proximity to each other, leading to contiguous groupings. The ■/■-dashed line highlights the fact that clusters, although sharing similar coloring for two of the three omics, can be separate due to the lack of measurements in the third omic type. The other colors indicate different trends in the respective clusters.*



(a) (b)

Figure 4: *Glyphs for different levels of detail using a blue-white-red diverging color scale (■■■■■) to encode measurement values. The high-detail version (a) shows discrete circle segments for each entity and the measurement coverage, while the low level-of-detail version (b) uses averaging to reduce the amount of information.*

After the cluster space is allocated, the nodes are distributed inside their cluster regions. For this final step in our layout pipeline, we use an adapted version of the ForceAtlas 2 [JVHB14] layout algorithm, modified to restrict the layout to a convex polygon. We randomly position, the pathway nodes inside the Voronoi-cell, disconnecting it from the UMAP embedding space. We apply weights using a simple weight function. Nodes belonging to the same Reactome superpathway will be weighed more than other edges (***R5***). Thus, all pathways in a cluster that belong to the same superpathway will be placed in closely together. This shows the semantic relations of the pathways based on the Reactome hierarchy and, thus, facilitates the analysis using biological domain knowledge. He et al. [HLY*19], in contrast, again only optimize the layout by using intra- and inter-cluster connectivity and minimizing edge crossings.

4.3. Pathway Glyphs

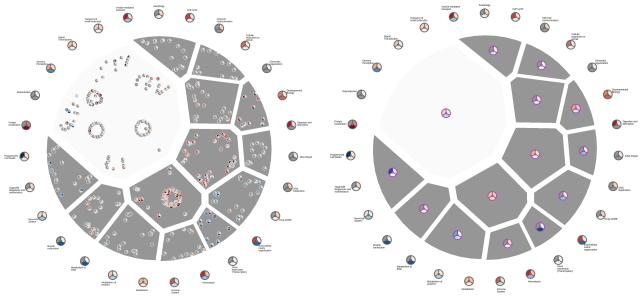
Since the nodes of the network represent whole pathways and multi-omics data is used, the amount of data per node can be large and diverse. This makes it challenging to give the user an overview of the measured data per pathway. We propose a radial glyph representing the multi-omics data contained in the respective pathway. Our glyph is divided into equally-sized radial sectors, depending on the amount of omics to be displayed, i.e., two half circles for two omics types, three third-circles for three omics types, etc. (***R1***/***R2***). Each sector is split into an outer and inner ring, as shown in Fig. 4 (right). The outer ring presents the regulation data per measured entity and is, therefore, again divided into equally-sized radial sub-sectors. Each sub-sector represents one omics measurement for an entity in the pathway (e.g., the transcription rate of a certain gene or the abundance of a certain protein). To convey this information in a simple way, we color the sub-segments according to their sorted measurement data, using a diverging color map from red (high positive regulation value) over white (neutral, no up-/down-regulation) to blue (high negative regulation value) (***R1***). The inner ring is used to display the number of measured molecular entities in relation to the absolute number of entities of the same type that are contained in the given pathway (dark gray portion). An annotated version of the glyph additionally shows these key information as text (Figure 9).

We also designed a simplified version of this glyph for scenarios in which the zoom level makes the detailed reading of values infeasible. The simplified glyph is shown in Fig. 4 (left). Here, the average of the regulation values for each omics type is computed and mapped to the corresponding color. Furthermore, we do not show the inner ring, thus providing more space for showing the average regulation value. This supports the *at-a-glance* analysis of the average regulation of the respective omics level per pathway (***R4***).

We chose a radial glyph for several reasons. Radial glyphs have been successfully applied in a similar scenario where a large amount of glyphs is drawn in proximity to each other [FFM12]. As nodes can be positioned rather close to each other in our layout, a linear glyph could lead to that undesired effect that bars could appear to be continuations of other, unrelated glyphs. Additionally, linear glyph designs could imply different degrees of importance to the different omics types due to the sorting. That is, radial glyphs allow for a clear separation of single pathways and do not impart any inherent hierarchy to the different omics (***R3***). Our circular design also easily scales with the number of displayed omcis types (***R2***).

5. Prototypical Application Design and Implementation Details

We implemented our approach as a prototypical web-based client-server application shown in Fig. 1. On the server side, we used a python flask [MRLU] backend to perform the mapping of the measurement variables to the Reactome pathways and to calculate the clustering using the UMAP [MHM20] and OPTICS [ABKS99] im-



(a) (b)

Figure 5: *When the zoom level is reduced below a threshold, the pathway nodes inside the clusters (a) are aggregated into larger nodes utilizing the low level-of-detail glyph (b).*

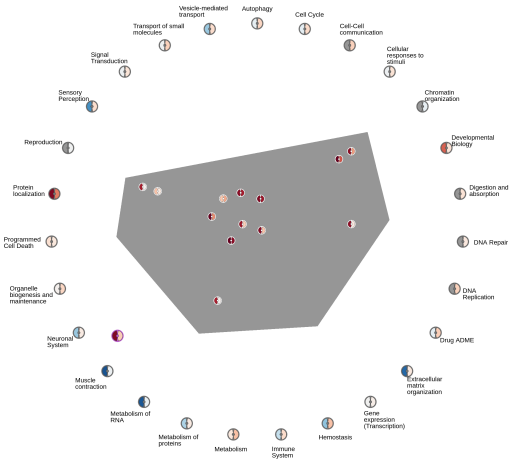


Figure 6: *Clicking a cluster aggregate node hides all structures not belonging to the cluster and instead displays the cluster in a zoomed state in the center of the canvas.*

plementations of scikit-learn [PVG*11]. Furthermore, all Reactome data like the detail pathways are served by the flask backend.

The remaining functionality of the prototype is facilitated by the client-side frontend, developed using TypeScript and Vue.js. For the network visualization, d3.js [BOH11], a Voronoi treemap plugin for d3 [Leb], and customized versions of the WebGL-accelerated Sigma.js [PJ] and ForceAtlas2 [JVHB14] are used. Our prototype contains four basic UI sections: an input sidebar for data selection, the main canvas showing the overview pathway network, a resizable view displaying the detail pathway, and a menu to parameterize the filtering and to control the intra-cluster layout. The detailed algorithmic pipeline can be found in the Supplement Material.

Since the visMOP network is designed to support the user in examining different data sets for different characteristics, there are various ways to customize the network to suit personal needs. The users can filter the input data by any data column supplied with the data. This can be, but is not limited to, the data variable used for mapping onto the Reactome hierarchy. In addition, the user can select from a list of statistical parameters per omics type, according to which the pathways should be clustered, and can, thus, influence the layout. After these two settings have been made by the user, the data is sent to the server to be summarized into pathway datasets. These are then sent back to the client together with the clustering data and subsequently, the network is drawn.

Depending on the zoom level, the pathway glyphs are either shown in full or reduced detail, as explained in Section 4.3. When the user further reduces the zoom level, the pathway nodes get aggregated into one node per cluster (Fig. 5). Each of these cluster nodes is shown using the low-detail variant of the proposed glyph design to show averages over all pathways that are part of the aggregated cluster. If the user selects a node, a focused and zoomed-in state of the respective cluster is shown, which again contains all pathway glyphs of this cluster (Fig. 6) (***R4***), i.e., hiding all other clusters.

As the size of the glyphs only allows for a limited in-detail inspection, the user can select one or more pathways to add them to a small multiples view, the pathway compare functionality, which allows the user to compare of different pathways (***R1***/***R4***). As shown in Fig. 11, the interactive glyphs additionally contain text labels showing the data values. This includes the total amount of entities, the proportion of measured entities, as well as the average measurements per omic when the corresponding section is hovered. The user can interact with these glyphs, switching their order via drag-and-drop or via mouse-over, which will displaying the name and associated data value of the respective molecular entity, or the average value of the respective omics category. Furthermore, utilizing the input data tables, the user can interactively search and select one or multiple molecular entities, highlighting pathways containing the union and intersection-sets of the selected entities, respectively (Fig. 7). The table (Fig. 1, ①) is located in a collapsible section in the left of the application, shared with the pathway compare functionality, between which it is possible to switch.

To reduce clutter and increase the focus on the selected characteristics, the user can apply filters, which act on the pathway level (see Fig. 1, ②). Here, several predefined filter criteria can be used to hide pathways in the visualization. In our prototype, these criteria are: belonging to a Reactome pathway superpathway, sum of measured values (relative and absolute), average changes, and relative or absolute coverage of the pathways for each omics type.

The overview visualization constitutes the centerpiece for this visual analytics approach. However, it has—by design—certain limitations with respect to the details that are available when zooming in. Due to the missing details about the interactions of the molecular entities, it can be difficult to investigate more complicated research questions or to further refine hypotheses generated via the overview visualization. Thus, users can select a pathway to display the pathway diagram on a molecular level (see Fig. 1, ⑤). We choose the layout and node/link encoding provided by the Reactome pathway browser (see Fig. 2 right) to preserve mental maps established by users familiar with Reactome (***R5***). We draw the diagram in gray scale and apply the same diverting red-white-blue color map that was used for the regulation values in the pathway glyphs to the molecular entities in the pathway diagram (see Fig. 8). That is, the omics measurements are color-coded onto the corresponding visual elements, which was shown to be beneficial in other tools [LSP*22;

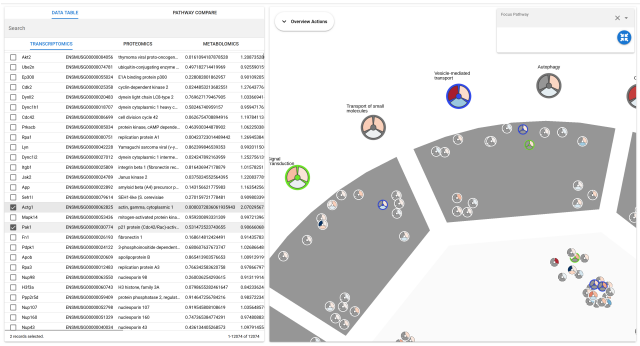


Figure 7: *Selecting entries in the input data table highlights pathways containing the selected entities in blue, indicating a union, or green, indicating an intersection of all selected entries.*

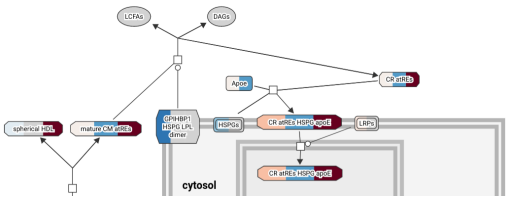


Figure 8: *Excerpt of a detail visualization of a pathway, using the Reactome layout. Measured entities are mapped onto their corresponding nodes. Metabolites are colored by the corresponding measurement value. Single protein nodes are split at the center, showing color information about transcriptomic and proteomic regulation. Complexes are divided in thirds and colored by transcriptomic, proteomic, and metabolomic regulation. These complex nodes can be clicked to show a labeled detail glyph containing all the members of the complex (analogous to the pathways, as shown in Fig. 9).*

LPB*17]. This way, users can investigate the regulation of the molecular entities in detail and immediately see where no data values are available. The pathway diagrams can also be panned and zoomed.

A central element of the Reactome detail pathway diagrams are complexes. This is due to the fact that many biological processes are not facilitated by single molecular entities, but instead through an interplay of several molecules or multi-meres of the same one. To

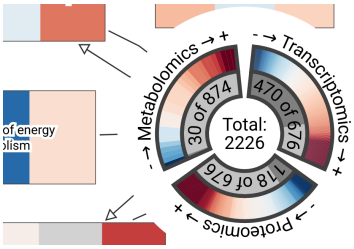


Figure 9: *Using the detail view of the "Metabolism" super pathway, an Overview glyph of the* Metabolism of lipids *sub-pathway can be shown by clicking on the corresponding pathway node.*

keep the familiar appearance of the original Reactome visualization and to reduce visual clutter, the complex elements are only colored by the average value per entity type (like the low-detail glyphs). Analogous to the overview glyphs, the user can click on a complex to show a glyph that shows each of the molecular entities partaking in the complex (Fig. 9).

6. Results and Discussion

In this section, we present an exemplary workflow using real-world multi-omics measurements, which was executed together with a domain expert, and the feedback provided by the expert. To this end, datasets were supplied by the expert that had previously been analyzed separately for the three different omics domains.

Our visMOP approach primarily focuses on generating a rich overview to facilitate an exploratory data analysis, while allowing more detailed observations using several linked detail-on-demand views and techniques. To our knowledge, none of the other methods and available tools offers a data-driven layout approach that is interactively parameterizable and allows expert users to generate network overviews that allow them to focus on exploring the data to discover entry points for further analysis. The recomputation of the layout only takes in the order of seconds for typical data set sizes. Consequently, users can adapt the layout interactively while investigating the data, in case the user want to explore the data with a different focus.

Below, we show the application of our prototypical implementation of visMOP on a real-life experimental datasets from *M. musculus* (house mouse) by performing a basic workflow to generate new hypotheses. The data [HKK*21] were acquired in an experimental study investigating the effects of high caloric feeding on mitochondrial lipid metabolism in the context of fatty liver disease. Samples were taken from liver tissue, and transcriptomic analysis was performed on whole tissue. Proteomics was performed on isolated mitochondria. Lipidomics analysis, i.e., a metabolomics analysis focused on lipids and lipid metabolites, was performed on both whole tissue and isolated mitochondria. This results in measurements for a reference control group and an interventional group, the group which was fed the high caloric feed. From these, differences per measurement, the fold change, can be calculated. From more than 12,000 and 900 measured transcripts and proteins, respectively, 5,936 and 654 entities could be mapped to Reactome identifiers. The lipidomics measurements were mapped to 37 identifiers, corresponding to either individual lipids or sums of lipid classes with mitochondrial or pan-cellular localization. While this represents a considerable loss of data, Reactome takes great care to only include interactions backed by literature evidence, increasing confidence in the overview [Jos04; VDS*07]. As new evidence is published, new interactions are added to Reactome with each new revision, reducing the loss of data.

For the initial clustering and layout, no input parameters were changed in visMOP to fully utilize the data, i.e., no filtering or setting the metrics used for clustering. The first inspection of the overview revealed that it was not focused enough for the planned investigation. Therefore, the input values that are considered in the clustering step of our layout pipeline were adapted: the minimum

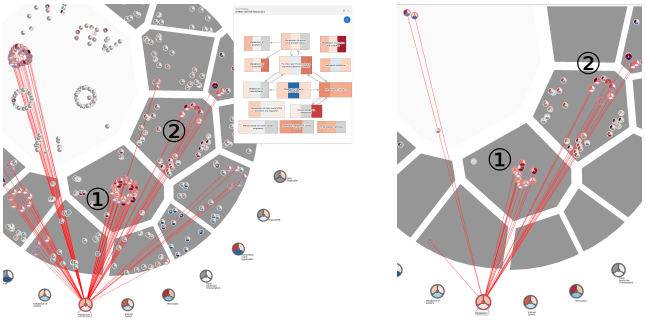


Figure 10: *Left: visMOP overview with the "Metabolism" superpathway selected, highlighting all pathways in the respective hierarchy, and showing the detail view of the "Metabolism" superpathway. Right: The visual filtering utility of the prototype was used to filtering pathways not containing metabolites*

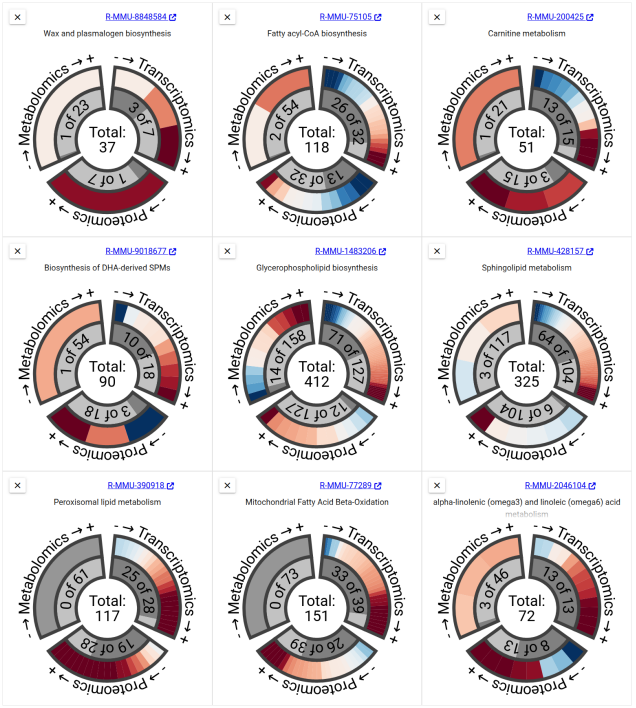


Figure 11: *By selecting pathways for the "Pathway Compare" utility of our prototype, high detail versions of the pathway glyphs can be easily compared. Here, pathways associated with lipid metabolism are shown for Cluster 1 (Fig. 10, ①)*

fold change limit for the metabolites was removed (since small changes in lipid concentration were of interest), unregulated entities were excluded from the clustering (since they are of little relevance here), and size of the Reactome pathways was not considered (since we did not want to focus on pathway size). After this adaption of the layout, the complete overview network was again inspected to search for any noticeable high-level trends (Fig. 3). In this case, a first general observation is that several clusters contain glyphs with a rather subdued coloring (Fig. 3 ■), indicating relatively little change in the omics. While other clusters show strong tendencies to red (Fig. 3 ■) indicating up-regulation, or mixed between red and blue (Fig. 3 ■) or blue and neutral (Fig. 3 ■). Another interesting observation is the similarity in transcriptome and proteome coloring of two cluster aggregate nodes (Fig. 3 ■/■-dashed), however, there is no metabolic data associated to the pathways in one of the clusters, which is a reason why they are separated. Furthermore, it is of notice that the aggregated cluster nodes in Fig. 3 ■ appear relatively similar. As the layout, however, uses data statics other than the average fold change, clusters might form which do have similar average fold change characteristics but differ otherwise.

As the data stem from an experiment designed to investigate the response to an energy-rich diet, an obvious starting point would be to inspect pathways that are a part of the "metabolism" superpathway (Fig. 10). To help to find agglomerations and patterns in these pathways, the intra-cluster layout positions pathways belonging to the same superpathway in proximity to each other and to the superpathway glyph. Selecting the latter highlights all pathways assigned to, in this case, the metabolomics superpathway. An immediate observation is that many of the associated pathways are located in the two clusters containing mostly red, i.e., up-regulated, pathways.

In agreement with the predominantly red color in the zoomed-out bird's eye view, the details of the metabolism superpathway glyph reveal that most of the transcripts, mitochondrial proteins and lipid metabolites detected and assigned to the pathway are increased in the liver of mice fed the energy-rich diet (Fig. 9). The inner sections of the glyph also show that the coverage is much higher for the transcripts than for proteins and metabolites, in agreement with the fact that these datasets were obtained from isolated mitochondria and using a lipidomics, i.e., a lipid-focused metabolomics analysis.

Next, the individual clusters connected to the *Metabolism* superpathway can be inspected. The bird's eye view shows that pathways pertaining to the metabolism superpathway form sub-clusters in multiple clusters and in the "noise"-cluster (Fig. 10 left). Among these sub-clusters, several do not contain data from all three omics types, in agreement with the preselection due to the mitochondria- and lipid-focused protein and metabolite analyzes—a fact that can be easily visualized using the filter functionality (Fig. 10 right).

One particularly large sub-cluster was selected for a detailed assessment of the individual pathways using the "compare pathways" functionality. Fig. 11 shows the pathways contained in this sub-cluster that are directly related to lipid metabolism. This view allows for a more detailed interpretation at the pathway level. In this example, one striking observation is that not only metabolites—i.e., lipids—but also transcripts and proteins are generally upregulated in pathways related to lipid metabolism. This includes several pathways concerned with the metabolism of fatty acids, in agreement with the original study [HKK*21] and indicates that the increased amount of fatty acids supplied with the fat-rich diet in the study caused an increased utilization of these in the abovementioned pathways. In contrast, individual proteins and transcripts assigned to the fatty acyl-CoA biosynthesis pathway, i.e., a pathway concerned with the synthesis of fatty acids, showed a higher degree of downregu-

lation. The metabolites associated with this pathway are colored in red, indicating an increase, but clicking this section of the icon quickly solves the apparent contradiction and reveals that these are fatty acids, likely derived from the diet and not from synthesis in the organism.

This concludes the exemplary workflow on how a hypothesis can be generated using visMOP, which can then later be confirmed via different means. In this case we showed a tendency in the data which was previously analyzed in detail and published [HKK*21]. There are a multitude of different entry points for further detail analysis, e.g., other clusters focusing on other processes. Additional analyses of the same dataset could, for example, investigate the makeup of the blue clusters, indicating downregulation, or the makeup of the with respect to the different omics differentially regulated cluster. This shows that the presented overview approach can be successfully used to open up the dataset to different research directions, increasing the knowledge-yield of time-consuming and costly experiments.

In general, the experts reported that our approach offers an innovative way of investigating the data. They particularly mentioned that the clustered view is helpful to get quick insights into the general structure of and an overview over the data. At the same time, they felt that visMOP's layout and design motivates to explore the data.

7. Conclusion and Outlook

We presented visMOP, a visual analytics approach for multi-omics pathways that combines pathway-level glyphs with a data-driven layout algorithm to facilitate an exploratory analysis workflow. Our main contribution is the overview visualization, not only abstracting the data but using the underlying structure to support the exploratory analysis. In our prototypical implementation, users can investigate experimental data from up to three different omics domains, to explore the measured data and search for interesting trends, correlations, and outliers. However, our approach is not theoretically limited to the presented omics types and could be extended straightforwardly to support more types. Similarly, while we currently use pathway data from Reactome, extending our approach to utilize other databases like KEGG would be straightforward. Filtering, details-on-demand, as well as level-of-detail visualizations allow the development of more focused hypotheses for further bioinformatics or biomedical analyses. We demonstrated the applicability of our approach in a basic workflow using a real-life data set from mice.

During the design process and the implementation of visMOP, several important insights were gained. In the beginning [KDC*20], we attempted to create a unified overview using existing pathway visualizations, like KEGG pathway maps, to show all details and fully maintain the user's mental map. This, however, resulted in a visualization that was too cluttered and difficult to use, as confirmed by our domain experts. We realized that combining familiar visualization, i.e., decorated detail pathways utilizing known layouts, with novel types of visualization yields a more accessible result. One such novel visualization is the data-driven clustered overview. We also realized that for complex and detailed data, a "fingerprint"-style [KO07] visualization, like our glyphs, can be more helpful than a more fine-grained display of the data. This is especially true for the presented overview concept, where general trends in the data might be more relevant than specific results, and could be transferred to similar works in which giving an overview is the main focus.

Current Limitations and Future Work: In its current form, visMOP only supports the direct comparison of one experimental group with a reference at one given time point. While this satisfies many experimental setups, some scenarios are not covered; including time-series experiments and multi-condition setups. These directions offer challenges, as the cluster space might change over time, making a consistent yet clutter-free visualization difficult.

While the glyph design allows an at the glance overview for the presented use cases, the radial division for the omic types might result in limited scalability. As the detail-level Reactome pathways themselves are limited in size, to be readily understandable by a human, this only affects the overview pathways. For the overview pathways, differentiating the entities can indeed become difficult, but in these cases, the outer channels act as a "fingerprint", allowing for a qualitative comparison of the overview pathways. Furthermore, while this is not an issue for three omics types, higher amounts of omics would make memorizing the order of omics more cumbersome. However, as most biomolecular experiments do not collect more than the omics types presented, with maybe one or two more, this has little impact on the applicability of the glyph.

Both hierarchical and direct connections between the different pathways are only drawn on demand. This is due to the high amount of edge crossings and the general high level of connectivity of the graph. While this preserves readability of the graph, it also makes it hard to evaluate connectivity trends of the graph as a whole, which contradicts the idea of an integrated, comprehensive data analysis we try to foster with visMOP. Thus, we plan to employ techniques like edge bundling to reduce the cluttering introduced when displaying the entirety of the edges while simultaneously making trends with regard to the edge connections more visible.

To increase the visual analytics capabilities of visMOPs, pathway enrichment analysis could be added. These enrichment values could either be used as further input values for the overview layout, or for filtering pathways. This could reduce the chance of including irrelevant pathways in the clustering or the visualization. Additionally, further aspects for analyzing the specific omics are conceivable. One example would be the integration of protein-protein interaction networks, which can be used to investigate if proteins of interest found during the previous analysis interact with each other.

While we already offer user choice by filtering of the input data and by means of selecting from a variety of derived variables, further possibilities to influence the clustering would increase its value. Weighting the derived variables, e.g., using UMAP with a different distance metric, would allow users to emphasize some attributes without disregarding others. Thus, investigating the effect of different distance metrics and even dimensionality reduction and clustering approaches on the quality of the results would be an interesting direction for future research.

Acknowledgments

We thank Jennifer Maurer for providing valuable feedback on the design and usability of visMOP. This work was supported by the

DFG/NSFC Sino-German mobility program (M-0257). Open Access funding enabled and organized by Projekt DEAL.

References

[ABKS99] ANKERST, MIHAEL, BREUNIG, MARKUS M., KRIEGEL, HANS-PETER, and SANDER, JÖRG. "OPTICS: ordering points to identify the clustering structure". *ACM SIGMOD Record* 28.2 (June 1999), 49–60. ISSN: 0163-5808. DOI: 10.1145/304181.304187 5, 6.

[AHN*21] ANDREA GUARRACINO, HEUMOS, SIMON, NAHNSEN, SVEN, et al. *ODGI: understanding pangenome graphs*. Tech. rep. Genomics Research Centre, Human Technopole, Milan, Italy, 2021, 2021.11.10.467921. DOI: 10.1101/2021.11.10.467921 3.

[ASA*22] ABDELAAL, MOATAZ, SCHIELE, NATHAN D., ANGERBAUER, KATRIN, et al. "Comparative Evaluation of Bipartite, Node-Link, and Matrix-Based Network Representations". *IEEE Transactions on Visualization and Computer Graphics* (2022), 1–11. ISSN: 1941-0506. DOI: 10.1109/TVCG.2022.3209427 3.

[Ass12] ASSENT, IRA. "Clustering high dimensional data". *WIREs Data Mining and Knowledge Discovery* 2.4 (July 2012), 340–350. ISSN: 1942-4787, 1942-4795. DOI: 10.1002/widm.1062 5.

[BBDW17] BECK, FABIAN, BURCH, MICHAEL, DIEHL, STEPHAN, and WEISKOPF, DANIEL. "A Taxonomy and Survey of Dynamic Graph Visualization: A Taxonomy and Survey of Dynamic Graph Visualization". *Computer Graphics Forum* 36.1 (Jan. 2017), 133–159. ISSN: 01677055. DOI: 10.1111/cgf.12791 2.

[BD05] BALZER, M. and DEUSSEN, O. "Voronoi Treemaps". *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* Minneapolis, MN, USA: IEEE, 2005, 49–56. ISBN: 978-0-7803-9464-3. DOI: 10.1109/INFVIS.2005.1532128 5.

[BM13] BREHMER, MATTHEW and MUNZNER, TAMARA. "A Multi-Level Typology of Abstract Visualization Tasks". *IEEE Transactions on Visualization and Computer Graphics* 19.12 (Dec. 2013). Conference Name: IEEE Transactions on Visualization and Computer Graphics, 2376–2385. ISSN: 1941-0506. DOI: 10/f5h3q4 4.

[BOH11] BOSTOCK, MICHAEL, OGIEVETSKY, VADIM, and HEER, JEFFREY. "D³ Data-Driven Documents". *IEEE Transactions on Visualization and Computer Graphics* 17.12 (Dec. 2011). Conference Name: IEEE Transactions on Visualization and Computer Graphics, 2301–2309. ISSN: 1941-0506. DOI: 10.1109/TVCG.2011.185 7.

[CAM19] CRUZ, ANTÓNIO, ARRAIS, JOEL P, and MACHADO, PENOUSAL. "Interactive and coordinated visualization approaches for biological data analysis". *Briefings in Bioinformatics* 20.4 (July 2019), 1513–1523. ISSN: 1467-5463, 1477-4054. DOI: 10.1093/bib/bby019 5.

[CMH*99] CHEN, DAGANG, MA, HAN, HONG, HENG, et al. "Regulation of Transcription by a Protein Methyltransferase". *Science* 284.5423 (June 1999), 2174–2177. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.284.5423.2174 2.

[CSB*20] CANZLER, SEBASTIAN, SCHOR, JANA, BUSCH, WIBKE, et al. "Prospects and challenges of multi-omics data integration in toxicology". *Archives of Toxicology* 94.2 (Feb. 2020), 371–388. ISSN: 0340-5761, 1432-0738. DOI: 10.1007/s00204-020-02656-y 2.

[CT12] CROSSWELL, LINDSEY C. and THORNTON, JANET M. "ELIXIR: a distributed infrastructure for European biological data". *Trends in Biotechnology* 30.5 (May 2012), 241–242. ISSN: 01677799. DOI: 10.1016/j.tibtech.2012.02.002 3.

[DGC*17] DUSSAUT, J S, GALLO, C A, CRAVERO, F, et al. "GeRNet: a gene regulatory network tool". *Biosystems.* 162 (Dec. 2017), 1–11. DOI: 10/gnhwwt 3.

[ENS*20] EIZENGA, JORDAN M., NOVAK, ADAM M., SIBBESEN, JONAS A., et al. "Pangenome Graphs". *Annual Review of Genomics and Human Genetics* 21.1 (2020). _eprint: https://doi.org/10.1146/annurev-genom-120219-080406, 139–162. DOI: 10/ghtrkn 3.

[FFM12] FISCHER, FABIAN, FUCHS, JOHANNES, and MANSMANN, FLORIAN. "ClockMap: Enhancing Circular Treemaps with Temporal Glyphs for Time-Series Data". *EuroVis - Short Papers* (2012), 97–101. DOI: 10.2312/PE/EUROVISSHORT/EUROVISSHORT2012/097-101 6.

[GJS*22] GILLESPIE, MARC, JASSAL, BIJAY, STEPHAN, RALF, et al. "The reactome pathway knowledgebase 2022". *Nucleic Acids Research* 50.D1 (Jan. 2022), D687–D692. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkab1028 2, 3.

[Her89] HERSHEY, J W. "Protein Phosphorylation Controls Translation Rates". *Journal of Biological Chemistry* 264.35 (1989), 20823–20826. ISSN: 00219258. DOI: 10.1016/S0021-9258(19)30005-5 2.

[HK99] HINNEBURG, ALEXANDER and KEIM, DANIEL A. "Optimal Grid-Clustering : Towards Breaking the Curse of Dimensionality in High-Dimensional Clustering". *VLDB '99: Proceedings of the 25th International Conference on Very Large Data Bases*. 1999, 506–517 5.

[HKK*21] HOENE, MIRIAM, KAPPLER, LISA, KOLLIPARA, LAXMIKANTH, et al. "Exercise prevents fatty liver by modifying the compensatory response of mitochondrial metabolism to excess substrate availability". *Molecular Metabolism* 54 (Dec. 2021), 101359. ISSN: 22128778. DOI: 10.1016/j.molmet.2021.101359 8–10.

[HLY*19] HE, SHENG, LIU, YI-JUN, YE, FEI-YUE, et al. "A new grid- and modularity-based layout algorithm for complex biological networks". *PLOS ONE* 14.8 (Aug. 2019). Ed. by LI, HUIJIA, e0221620. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0221620 5, 6.

[HMB*18] HADFIELD, JAMES, MEGILL, COLIN, BELL, SIDNEY M, et al. "Nextstrain: real-time tracking of pathogen evolution". *Bioinformatics* 34.23 (2018), 4121–4123. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bty407 3.

[JKS06] JUNKER, BJÖRN H, KLUKAS, CHRISTIAN, and SCHREIBER, FALK. "VANTED: A system for advanced data analysis and visualization in the context of biological networks". *BMC Bioinformatics* 7.1 (Dec. 2006), 109. ISSN: 1471-2105. DOI: 10.1186/1471-2105-7-109 4.

[Jos04] JOSHI-TOPE, G. "Reactome: a knowledgebase of biological pathways". *Nucleic Acids Research* 33.Database issue (Dec. 2004), D428–D432. ISSN: 1362-4962. DOI: 10.1093/nar/gki072 8.

[JVHB14] JACOMY, MATHIEU, VENTURINI, TOMMASO, HEYMANN, SEBASTIEN, and BASTIAN, MATHIEU. "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software". *PLOS ONE* 9.6 (Oct. 2014). Publisher: Public Library of Science, e98679. ISSN: 1932-6203. DOI: 10/f3mn4z 6, 7.

[KBC*19] KARP, PETER D., BILLINGTON, RICHARD, CASPI, RON, et al. "The BioCyc collection of microbial genomes and metabolic pathways". *Briefings in Bioinformatics* 20.4 (July 2019), 1085–1093. ISSN: 1477-4054. DOI: 10.1093/bib/bbx085 3.

[KDC*20] KRONE, MICHAEL, DRÄGER, ANDREAS, COBANOGLU, EBRU, et al. "A Web-based Visual Analytics Application for Biological Networks." *EuroVis (Posters)*. 2020, 41–43. DOI: 10.2312/eurp20201124 4, 10.

[KEC06] KELLER, RENÉ, ECKERT, CLAUDIA M., and CLARKSON, P. JOHN. "Matrices or Node-Link Diagrams: Which Visual Representation is Better for Visualising Connectivity Models?": *Information Visualization* 5.1 (Mar. 2006), 62–76. ISSN: 1473-8716, 1473-8724. DOI: 10.1057/palgrave.ivs.9500116 3.

[KG00] KANEHISA, MINORU and GOTO, SUSUMU. "KEGG: Kyoto Encyclopedia of Genes and Genomes". *Nucleic Acids Research* 28.1 (2000), 27–30. ISSN: 0305-1048 3.

[KKPP20] KOUTROULI, MIKAELA, KARATZAS, EVANGELOS, PAEZ-ESPINO, DAVID, and PAVLOPOULOS, GEORGIOS A. "A Guide to Conquer the Biological Network Era Using Graph Theory". *Front. Bioeng. Biotechnol.* 8 (2020), 34. DOI: 10.3389/fbioe.2020.00034 2, 3.

[KO07] KEIM, DANIEL A. and OELKE, DANIELA. "Literature Fingerprinting: A New Method for Visual Literary Analysis". *2007 IEEE Symposium on Visual Analytics Science and Technology*. Sacramento, CA, USA: IEEE, Oct. 2007, 115–122. ISBN: 978-1-4244-1659-2. DOI: 10.1109/VAST.2007.4389004 10.

[KS08] KARLEBACH, GUY and SHAMIR, RON. "Modelling and analysis of gene regulatory networks". *Nature Reviews. Molecular Cell Biology* 9.10 (Oct. 2008), 770–780. ISSN: 1471-0080. DOI: 10.1038/nrm2503 3.

[Lad06] LADURNER, ANDREAS G. "Rheostat Control of Gene Expression by Metabolites". *Molecular Cell* 24.1 (Oct. 2006), 1–11. ISSN: 10972765. DOI: 10.1016/j.molcel.2006.09.002 2.

[Lat13] LATCHMAN, D.S. "Eukaryotic Transcription Factors". *Brenner's Encyclopedia of Genetics*. Elsevier, 2013, 537–539. ISBN: 978-0-08-096156-9. DOI: 10.1016/B978-0-12-374984-0.01548-5 2.

[Leb] LEBEAU, FRANCK. *d3-voronoi-map*. https://github.com/Kcnarf/d3-voronoi-map (last accessed 17.3.2023) 7.

[LPB*17] LUO, WEIJUN, PANT, GAURAV, BHAVNASI, YESHVANT K, et al. "Pathview Web: user friendly pathway visualization and data integration". *Nucleic acids research* 45.W1 (2017). Publisher: Oxford University Press, W501–W508. DOI: 10.1093/nar/gkx372 4, 8.

[LSP*22] LIU, TIANYUAN, SALGUERO, PEDRO, PETEK, MARKO, et al. "PaintOmics 4: new tools for the integrative analysis of multi-omics datasets supported by multiple pathway databases". *Nucleic Acids Research* 50.W1 (July 2022), W551–W559. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkac352 4, 7.

[MGM*19] MCGEE, F., GHONIEM, M., MELANÇON, G., et al. "The State of the Art in Multilayer Network Visualization". *Computer Graphics Forum* 38.6 (Sept. 2019), 125–149. ISSN: 0167-7055, 1467-8659. DOI: 10.1111/cgf.13610 2, 4.

[MHM20] MCINNES, LELAND, HEALY, JOHN, and MELVILLE, JAMES. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction". *arXiv:1802.03426 [cs, stat]* (2020). URL: http://arxiv.org/abs/1802.03426 (visited on 04/11/2023) 5, 6.

[Mic17] MICHAL, GERHARD. *Roche - Biochemical Pathways*. 2017. URL: http://biochemical-pathways.com/ 3.

[MMF17] MURRAY, PAUL, MCGEE, FINTAN, and FORBES, ANGUS G. "A taxonomy of visualization tasks for the analysis of biological pathway data". *BMC bioinformatics* 18.2 (2017). Publisher: BioMed Central, 1–13. DOI: 10.1186/s12859-016-1443-5 4.

[MRLU] MÖNNICH, ADRIAN, RONACHER, ARMIN, LORD, DAVID, and UNTERWADITZER, MARKUS. *Flask*. https://palletsprojects.com/p/flask/ (last accessed 2.12.2022) 6.

[Mun15] MUNZNER, TAMARA. *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press, 2015. ISBN: 978-1-4987-5971-7 4.

[NB12] NOCAJ, ARLIND and BRANDES, ULRIK. "Computing Voronoi Treemaps: Faster, Simpler, and Resolution-independent". *Computer Graphics Forum* 31.3pt1 (June 2012), 855–864. ISSN: 01677055. DOI: 10.1111/j.1467-8659.2012.03078.x 5.

[NDG*17] NORONHA, ALBERTO, DANÍELSDÓTTIR, ANNA DRÖFN, GAWRON, PIOTR, et al. "ReconMap: an interactive visualization of human metabolism". *Bioinformatics* 33.4 (2017), 605–607. DOI: 10.1093/bioinformatics/btw667 3.

[NMSL19] NOBRE, C., MEYER, M., STREIT, M., and LEX, A. "The State of the Art in Visualizing Multivariate Networks". *Computer Graphics Forum* 38.3 (2019), 807–832. ISSN: 1467-8659. DOI: 10.1111/cgf.13728 2.

[NYP12] NEPUSZ, TAMÁS, YU, HAIYUAN, and PACCANARO, ALBERTO. "Detecting overlapping protein complexes in protein-protein interaction networks". *Nature Methods* 9.5 (Mar. 2012), 471–472. ISSN: 1548-7105. DOI: 10.1038/nmeth.1938 3.

[PCZ*21] PANG, ZHIQIANG, CHONG, JASMINE, ZHOU, GUANGYAN, et al. "MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights". *Nucleic Acids Research* 49.W1 (July 2021), W388–W396. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkab382 4.

[PDA10] PRICE, MORGAN N., DEHAL, PARAMVIR S., and ARKIN, ADAM P. "FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments". *PLOS ONE* 5.3 (2010), e9490. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0009490 3.

[Pes14] PESSOA, LUIZ. "Understanding brain networks and brain organization". *Physics of Life Reviews* 11.3 (Sept. 2014), 400–435. ISSN: 15710645. DOI: 10.1016/j.plrev.2014.03.005 3.

[PJ] PLIQUE, GUILLAUME and JACOMY, ALEXIS. *Sigma.js*. https://www.sigmajs.org/ (last accessed 2.12.2022) 7.

[PVG*11] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., et al. "Scikit-learn: Machine Learning in Python". *Journal of Machine Learning Research* 12 (2011), 2825–2830 7.

[RJH*12] ROHN, HENDRIK, JUNKER, ASTRID, HARTMANN, ANJA, et al. "VANTED v2: a framework for systems biology applications". *BMC systems biology* 6.1 (2012). Publisher: Springer, 1–13. DOI: 10.1186/1752-0509-6-139 4.

[RKK*19] RAUDVERE, UKU, KOLBERG, LIIS, KUZMIN, IVAN, et al. "g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update)". *Nucleic Acids Research* 47.W1 (July 2019), W191–W198. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkz369 4.

[Rou87] ROUSSEEUW, PETER J. "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". *Journal of Computational and Applied Mathematics* 20 (Nov. 1987), 53–65. ISSN: 03770427. DOI: 10.1016/0377-0427(87)90125-7 5.

[RS18] RAPPOPORT, NIMROD and SHAMIR, RON. "Multi-omic and multi-view clustering algorithms: review and cancer benchmark". *Nucleic Acids Research* 46.20 (Nov. 2018), 10546–10562. ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gky889 2.

[SM03] SPIRIN, VICTOR and MIRNY, LEONID A. "Protein complexes and functional modules in molecular networks". *Proceedings of the National Academy of Sciences of the United States of America* 100.21 (Oct. 2003), 12123–12128. ISSN: 0027-8424. DOI: 10.1073/pnas.2032324100 3.

[SPK*19] SCHWACKE, RAINER, PONCE-SOTO, GABRIEL Y., KRAUSE, KIRSTEN, et al. "MapMan4: A Refined Protein Classification and Annotation Framework Applicable to Multi-Omics Data Analysis". *Molecular Plant* 12.6 (2019), 879–892. DOI: 10.1016/j.molp.2019.01.003 4.

[VDS*07] VASTRIK, IMRE, D'EUSTACHIO, PETER, SCHMIDT, ESTHER, et al. "Reactome: a knowledge base of biologic pathways and processes". *Genome Biology* 8.3 (2007), R39. ISSN: 14656906. DOI: 10.1186/gb-2007-8-3-r39 8.

[vLKS*11] Von LANDESBERGER, T., KUIJPER, A., SCHRECK, T., et al. "Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges". *Computer Graphics Forum* 30.6 (2011), 1719–1749. ISSN: 1467-8659. DOI: 10.1111/j.1467-8659.2011.01898.x 2.

[WMWH15] WEGNER, ANDRE, MEISER, JOHANNES, WEINDL, DANIEL, and HILLER, KARSTEN. "How metabolites modulate metabolic flux". *Current Opinion in Biotechnology* 34 (Aug. 2015), 16–22. ISSN: 09581669. DOI: 10.1016/j.copbio.2014.11.008 2.

[WNSV19] WU, HSIANG-YUN, NÖLLENBURG, MARTIN, SOUSA, FILIPA L, and VIOLA, IVAN. "Metabopolis: scalable network layout for biological pathway diagrams in urban map style". *BMC Bioinformatics* 20.1 (2019), 1–20. DOI: 10.1186/s12859-019-2779-4 3.

[WZR*19] WU, CEN, ZHOU, FEI, REN, JIE, et al. "A Selective Review of Multi-Level Omics Data Integration Using Variable Selection". *High-Throughput* 8.1 (2019), 4. DOI: 10.3390/ht8010004 2.

[YSD*17] YANG, XINSONG, SHI, LEI, DAIANU, MADELAINE, et al. "Blockwise Human Brain Network Visual Comparison Using NodeTrix Representation". *IEEE transactions on visualization and computer graphics* 23.1 (Jan. 2017), 181–190. ISSN: 1941-0506. DOI: 10.1109/TVCG.2016.2598472 3.

[ZX18] ZHOU, GUANGYAN and XIA, JIANGUO. "OmicsNet: a web-based tool for creation and visual analysis of biological networks in 3D space". *Nucleic acids research* 46.W1 (2018). Publisher: Oxford University Press, W514–W522. DOI: 10.1093/nar/gky510 4.