MOBS - Multi-Omics Brush for Subgraph visualisation

D.Heylen[†,1,2], J. Peeters[†,1], G. Ertaylan[2], J. Hooyberghs[2,3], and J.Aerts[1],

[1] Data Science Institute (DSI), CENSTAT, Hasselt University, Diepenbeek, Belgium.
[2] Flemish institute for Technological Research (VITO) Mol, Belgium.
[3] Theoretical Physics, Data Science Institute (DSI), Hasselt University, Diepenbeek, Belgium.

Abstract

One of the big opportunities in multi-omics analysis is the identification of interactions between molecular entities and their association with diseases. In analyzing and expressing these interactions in the search for new hypotheses, multi-omics data is often either translated into matrices containing pairwise correlations and distances, or visualized as node-link diagrams. A major problem when visualizing large networks however is the occurrence of hairball-like graphs, from which little to none information can be extracted. It is of interest to investigate subgroups of markers that are closely associated with each other, rather than just looking at the overload of all interactions. Hence, we propose MOBS (Multi-Omics Brush for Subgraph visualisation), a web-based visualisation interface that can provide both an overview and detailed views on the data. By means of a two dimensional brush on a heatmap that includes hierarchical clustering, relationships of interest can be extracted from a fully connected graph, to enable detailed analysis of the subgraph of interest.

CCS Concepts

• ***Human-centered computing*** → *Visualization toolkits; Graph drawings;*

1. Introduction

The rapidly expanding amount of multi-level (omics) data requires concerted efforts on data integration. At this point, graph theory and biological networks are widely applied to represent multi-omics data [Heo21], but it remains a challenge to use these in a way so that a compartmentalized focus can be generated within a holistic view of (disease-related) molecular biology [CBB*20]. A user interface that can extract interesting subgraphs out of an overload of interactions has the potential to address this challenge [PSBB*21, KDSM20]. As such, enabling comprehensive, detailed exploratory analysis, that has the ability to generate hypotheses on bio-molecular interaction patterns hidden in hairball-like graphs.

Therefore, MOBS (Multi-Omics Brush for Subgraph visualisation) is developed as a solution. MOBS provides an overview of all relationships within a heatmap visualisation of an adjacency matrix, complemented with a node-link diagram to provide a more detailed view on the topological structure. In practice, the adjacency matrix would serve as a canvas to brush on, resulting in only that specific region of interest to be shown in the node-link diagram, from the otherwise hairball-like visual (figure 1).

Due to the tool's generic input requirements, different types of data can be visualized with the interface. Every matrix that stores relationships between its rows and columns can be transposed into a network structure (e.g. correlation-, distance-, or interaction matrix). Furthermore, based on the principle that every data has an underlying shape, the topology of a data set (shape) can be extracted using Topological Data Analysis (TDA) (e.g. STAD [AA21]). The result will be a network that can be used to explore the data, consisting of links between all similar data points [CM21]. Hence, multi-omics data can be visually explored by MOBS using multiple points of views, creating the potential for hypothesis generation.

Our software is available at `https://github.com/driesheylen123/Multi_omics_exploration`. The GitHub repository contains some example data and R scripts used to generate the images in this poster. A screenshot of the entire toolset is shown in the appendix.

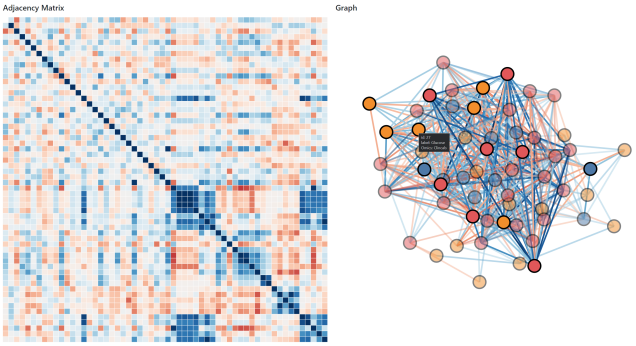
2. Related Work

Regarding the relevant state of the art for analyses and exploration of multiple omics data, a distinction can be made between *'computational integration frameworks'* and *'visual (network) explorations'* endeavors. MOFA [AVA*18] and COMBI [Haw20] are currently acknowledged as valuable computational frameworks for unsupervised discovery of the principal axes in multiple omics parameters from the same samples. However, these analyses are more rigid and not always suited for exploration purposes.

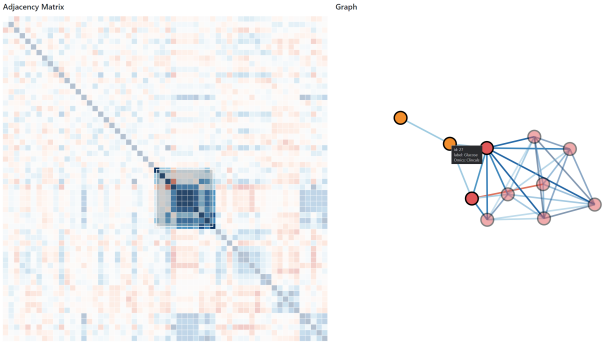
In visualisation methodology, several actions have been taken to tackle the problem of hairball-like graphs. Edge et al. (2018)

† D. Heylen and J. Peeters contributed equally to this work.

DOI: 10.2312/evp.20221116



(a) *Fully connected graph*



(b) *Brushed subgraph*

Figure 1: *Exploratory brushing of a biological correlation network. (a) The adjacency matrix shows the original Pearson correlation between metabolites (top,left) and clinical (bottom,right) parameters across a number of individuals. The edges in the corresponding node-link diagram display all pairwise correlations with an absolute value above 0.3. (b) A sub-graph is shown as a node-link diagram after applying the rectangular brush on the heatmap. This interaction leads to the interesting finding that the clinical parameter glucose lays a connection with a metabolite cluster, a finding that could not be observed with the fully connected graph.*

[ELMW18] categorize the current efforts into *'interactive'* and *'algorithmic approaches'*, focused on either *'graph aggregation'* or *'graph filtering'*. Graph aggregation makes use of common attributes to merge nodes and edges to create a simplified structure (e.g. NodeTrix [HFM07], or MCLEAN [AA18]), whereas graph filtering implies removal of nodes and edges based on heuristics or sampling (e.g. Responsive matrix cells [HBS*20]). MOBS adds to the interactive approaches, following similar design principles as in MatrixExplorer (visualiasation software developed for the analysis of social networks) [HF06], by using the adjacency matrix as a brush for graph filtering.

3. Implementation

MOBS is a web-application developed in Svelte (`https://svelte.dev/`), using the D3.js library [BOH11], and can be either installed and run locally or be used from the online hosted version on vercel (`https://mobs.vercel.app/`). The source code is publicly available, and detailed instructions on how to use and install the tool are provided in the README.md file in the GitHub repository, as well as the complementary data and scripts generated within the use case to demonstrate and evaluate the functionality of the tool.

3.1. Input requirements

Data can be uploaded locally in JSON format, containing a node and link list. An example of the required file structure, as well as R and Python scripts to generate the input format are provided in the README.md file in the GitHub repository. Additional variables such as node or edge values can be added, and will appear in the corresponding colouring options.

3.2. User Interface

MOBS contains various functionalities that enable efficient analyses. Different styling options are available for both the adjacency matrix and the node-link diagram (e.g. adjusting color and size of nodes and edges based on node values, edge weights, clusters, or any other variable included in the data). Tooltips are available in the node-link diagrams displaying information on all variables provided in the node list, and edge weights. In addition, node highlighting and dragging is included for easy reference of interactions.

The tool comes with several interactive features to optimize the user experience. The key interactive feature of the tool resolves around brushing on the adjacency matrix to extract sub parts of the graph. One can also zoom on both the adjacency matrix and node-link diagram (triggered by holding shift key in combination with mouse events) to obtain more detail. In addition, a threshold can be set to filter links in the node-link diagram based on edge weight. Since the identification of interesting (multi-omics) clusters is crucial to identify parameters with similar interaction patterns, hierarchical clustering is included in the tool using the DRUIDjs [CKS20] library to rearrange the order of the nodes across all data types in the adjacency matrix. Whenever a clustering method is chosen, the corresponding dendrogram will be added on top of the heatmap. A clustering threshold can be set to define the maximum depth of clusters, nodes can be assigned to.

4. Conclusion and Future Work

With MOBS we propose a tool that enables the expert user to include and explore different types of omics data. Scalability is the main focus for future work. This toolset provides a clear visual overview of up to 500 parameters, but above that, the computational load to render the visualisations becomes too high. Hence, additional actions need to be taken to either improve the rendering of larger data files (e.g. use WebGL or Canvas based rendering libraries), or include algorithms that rely on clustering in the preliminary network generation process (e.g. Mapper [SMC07]).

5. Acknowledgements

D.H. and J.P. are funded through Hasselt University BOF grants (BOF20OWB29 & BOF20OWB33). D.H. also receives funding from VITO (R-11362) .

References

[AA18] ALCAIDE D., AERTS J.: Mclean: Multilevel clustering exploration as network. *PeerJ Computer Science 4* (Jan. 2018), e145. URL: https://doi.org/10.7717/peerj-cs.145, doi:10.7717/peerj-cs.145. 2

[AA21] ALCAIDE D., AERTS J.: Spanning trees as approximation of data structures. *IEEE Transactions on Visualization and Computer Graphics 27*, 10 (2021), 3994–4008. doi:10.1109/TVCG.2020.2995465. 1

[AVA*18] ARGELAGUET R., VELTEN B., ARNOL D., DIETRICH S., ZENZ T., MARIONI J. C., BUETTNER F., HUBER W., STEGLE O.: Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology 14*, 6 (jun 2018). doi:10.15252/MSB.20178124. 1

[BOH11] BOSTOCK M., OGIEVETSKY V., HEER J.: D^3 data-driven documents. *IEEE transactions on visualization and computer graphics 17*, 12 (2011), 2301–2309. doi:10.1109/TVCG.2011.185. 2

[CBB*20] COMTE B., BAUMBACH J., BENIS A., BASÍLIO J., DEBELJAK N., FLOBAK Å., FRANKEN C., HAREL N., HE F., KUIPER M., ET AL.: Network and systems medicine: position paper of the european collaboration on science and technology action on open multiscale systems medicine. *Network and Systems Medicine 3*, 1 (2020), 67–90. doi:10.1089/nsm.2020.0004. 1

[CKS20] CUTURA R., KRALJ C., SEDLMAIR M.: DRUIDJS — A JavaScript Library for Dimensionality Reduction. In *2020 IEEE Visualization Conference (VIS)* (2020), IEEE, pp. 111–115. doi:10.1109/VIS47514.2020.00029. 2

[CM21] CHAZAL F., MICHEL B.: An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence 4* (2021). doi:10.3389/frai.2021.667963. 1

[ELMW18] EDGE D., LARSON J., MOBIUS M., WHITE C.: Trimming the hairball: Edge cutting strategies for making dense graphs usable. In *2018 IEEE International Conference on Big Data (Big Data)* (2018), IEEE, pp. 3951–3958. doi:10.1109/BigData.2018.8622521. 2

[Haw20] Model-based joint visualization of multiple compositional omics datasets. *NAR Genomics and Bioinformatics 2*, 3 (2020). doi:10.1093/nargab/lqaa050. 1

[HBS*20] HORAK T., BERGER P., SCHUMANN H., DACHSELT R., TOMINSKI C.: Responsive matrix cells: A focus+ context approach for exploring and editing multivariate graphs. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2020), 1644–1654. doi:10.1109/TVCG.2020.3030371. 2

[Heo21] Integrative multi-omics approaches in cancer research: From biological networks to clinical subtypes. *Molecules and Cells 44*, 7 (2021), 433–443. doi:10.14348/molcells.2021.0042. 1

[HF06] HENRY N., FEKETE J.-D.: Matrixexplorer: a dual-representation system to explore social networks. *IEEE transactions on visualization and computer graphics 12*, 5 (2006), 677–684. doi:10.1109/TVCG.2006.160. 2

[HFM07] HENRY N., FEKETE J.-D., MCGUFFIN M. J.: Nodetrix: a hybrid visualization of social networks. *IEEE transactions on visualization and computer graphics 13*, 6 (2007), 1302–1309. doi:10.1109/TVCG.2007.70582. 2

[KDSM20] KRASSOWSKI M., DAS V., SAHU S. K., MISRA B. B.: State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing. *Frontiers in Genetics 11* (dec 2020), 1598. doi:10.3389/FGENE.2020.610798/BIBTEX. 1

[PSBB*21] PICARD M., SCOTT-BOYER M. P., BODEIN A., PÉRIN O., DROIT A.: Integration strategies of multi-omics data for machine learning analysis. *Computational and Structural Biotechnology Journal 19* (jan 2021), 3735–3746. URL: https://doi.org/10.1016/j.csbj.2021.06.030, doi:10.1016/J.CSBJ.2021.06.030. 1

[SMC07] SINGH G., MÉMOLI F., CARLSSON G.: Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. *Eurographics Symposium on Point-Based Graphics* (2007). doi:10.2312/SPBG/SPBG07/091-100. 2