Data-dependent visualization of biological networks in the web-browser with NDExEdit

Florian Auer[1*], Simone Mayer[1], Frank Kramer[1]

1 Department of IT-Infrastructure for Translational Medical Research, Faculty of Applied Computer Science, University of Augsburg, Augsburg, Germany

* florian.auer@informatik.uni-augsburg.de

Abstract

Networks are a common methodology used to capture increasingly complex associations between biological entities. They serve as a resource of biological knowledge for bioinformatics analyses, and also comprise the subsequent results. However, the interpretation of biological networks is challenging and requires suitable visualizations dependent on the contained information. The most prominent software in the field for the visualization of biological networks is Cytoscape, a desktop modeling environment also including many features for analysis.

A further challenge when working with networks is their distribution. Within a typical collaborative workflow, even slight changes of the network data force one to repeat the visualization step as well. Also, just minor adjustments to the visual representation not only need the networks to be transferred back and forth. Collaboration on the same resources requires specific infrastructure to avoid redundancies, or worse, the corruption of the data. A well-established solution is provided by the NDEx platform where users can upload a network, share it with selected colleagues or make it publicly available.

NDExEdit is a web-based application where simple changes can be made to biological networks within the browser, and which does not require installation. With our tool, plain networks can be enhanced easily for further usage in presentations and publications. Since the network data is only stored locally within the web browser, users can edit their private networks without concerns of unintentional publication. The web tool is designed to conform to the Cytoscape Exchange (CX) format as a data model, which is used for the data transmission by both tools, Cytoscape and NDEx. Therefore the modified network can be exported as a compatible CX file, additionally to standard image formats like PNG and JPEG.

Author summary

Relations in biological research are often visualized as networks. For instance, if two proteins interact with each other during a certain process, the corresponding network would show two nodes connected by one edge. But the fact that the interaction between the two exists, may not be enough. With established software solutions like Cytoscape we can add all the information we have about our nodes and their interaction to our data foundation. Furthermore, we can change the visual appearance of our nodes and their interaction based on this information.

For example, if our network contains 20 nodes, that all interact with each other, but the strength of these interactions each range between 0 and 1, we can illustrate that by

making the edges wider for strong interactions and slimmer for weak interactions. Thus, our visualization is enriched with valuable information. As of now these data-dependent modifications can only be made with a desktop client.

We introduce NDExEdit, a web-based solution for visualization changes to networks that conform to the CX data format. It allows us to import networks directly from the NDEx platform and apply changes to the visualization – including all types of mappings, one of which was briefly described above.

Introduction

Networks are well-established in a wide range of fields in biology [1–3], and are often used, either as a source or result, in biological research. Information associated with the individual nodes or edges can go far beyond name and type, thus increasing its complexity. Within common bioinformatics workflows data integration, network analysis, and visualization accompany each other [4,5], and comprise fundamental challenges of combining various tools.

The information-rich data contained in biological networks provide the opportunity for comprehensive visualization but requires powerful tools to achieve. Cytoscape [6] is the most prominent desktop software for biological network analysis and visualization. It employs a data-dependent visualization strategy by applying so-called "attribute-to-visual-mappings", where a node's or edge's attribute translates to its visual representation. Besides its support for large networks and its rich set of features, Cytoscape comes with overhead for quick results and a steep learning curve.

A major challenge when working with networks is their distribution. Collaboration on the same resources requires specific infrastructure to avoid redundancies, or worse, the corruption of the data. A well-established solution is provided by the NDEx platform [7,8] where users can upload a network, share it with selected colleagues or make it publicly available. NDEx also holds the feature to provide your private networks solely to the reviewers of a submitted paper, to protect the data until publication.

NDEx is tightly connected to Cytoscape, which reveals itself in the mutual integration of both platforms. For the transmission of the networks the Cytoscape Exchange (CX) data structure [9] was developed, which not only includes the structural information of the networks but also instructions for its visual representation.

There is a recent trend in software development towards web-based solutions. Desktop applications require individual installations, which is not possible in all cases for various reasons and also brings further expense for maintenance. Furthermore, accessibility across different devices grows in importance, while web-based applications provide secure access to centralized data. In the following, we illustrate how our lightweight web application NDExEdit implements current web technologies and thereby facilitates the data-dependent visualization of biological networks.

Methods

Network data model

CX is a JSON (JavaScript Object Notation) based data structure designed for the transmission of biological networks between web applications and servers. The different types of information within a network are organized into single aspects of the network. These modular components separate the basic network structure from additional information and thus enable to only load the parts of the network that are of interest for an application. Since CX is designed as a transmission format, this reduces the amount of data needed to be transferred, but still combines all data as one coherent network.

The aspects have a defined scheme for the elements they can contain that must be followed. This includes definitions for core aspects, concerning the network topology and attributes, and aspects contributed by Cytoscape handling the visual representation. They link to each other by referencing the internal ID used in the aspects, for example, refer edges the IDs of the nodes aspects they are connecting. Furthermore, it is possible to include own custom aspects without a strict definition, that will be stored at the NDEx platform, but not processed or validated.

Implementation details

The client-side visualization of networks is realized using Cytoscape.js [10]. It is a JavaScript library for browser and server-based graph rendering, including layout algorithms for positioning nodes. One of its key features is the separation of data and its representation: stylesheets are used to data-dependently select network elements and assign visual properties to them.

Cytoscape.js does not natively support the handling of networks in CX format but is used in the front-end of the NDEx platform to visualize the CX networks. Their mapping script was incorporated into NDExEdit to assure a consistent visual representation in all software tools, including Cytoscape. Therefore, modifications of the script were necessary to enable highlighting and export of the networks.

The functionality of NDExEdit rests upon the Angular [11] platform, an open-source framework for building single-page web applications. It follows the Model-View-Controller (MVC) design pattern which reduces the code required for implementing the web application. Angular is based on TypeScript [12] as the programming language, which brings advantages for development in form of static typing and support of class-based object-oriented programming (OOP).

The layout of the web application is realized using the Bootstrap [13] framework. It is an open-source CSS framework for front-end development, containing design templates for interface components.

Results

Web-application

Here, the user can either load a network directly from NDEx into the app or upload a local CX file, which both become accessible in the list. The home button of any subordinate page leads back to this page to be able to switch between networks.

By default, the breast cancer protein-protein interaction network by Minkyu Kim [14] is provided for demonstration purposes. The network contains the interactome of all high-confidence PPIs detected across the three breast cell lines MCF7, MDA-MB-231, and MCF10A. Besides the valuable information contained in this network, it is also a great example of how the visual representation (Fig 1) supports the comprehension of the underlying data. Therefore, it will be used in the following to demonstrate the capabilities of NDExEdit to define and edit the attribute mappings dependent on the network data.

When accessing a network in NDExEdit, general information about it will be shown next to its visualization. This view can be customized by toggling the sides or moving the separating border in any direction. The general information panel provides an overview of all node, edge, and network attributes of the network. While the network attributes can be edited directly, the remaining attributes can be explored for their distribution (Fig 4A) and the coverage of the nodes and edges by this attribute.

Additionally, the network can be inspected by creating rules on the values of the node and edge attributes to be highlighted in the graph.

The visualization of the network is interactive, which means that it can be zoomed and shifted, and also the nodes and edges can be selected and moved. Detailed information about the selected elements appears on top in the information panel to be able to compare its content. With the available buttons, the graph can be fit to the viewport and for better overview and performance improvements, the labels in the network can be hidden.
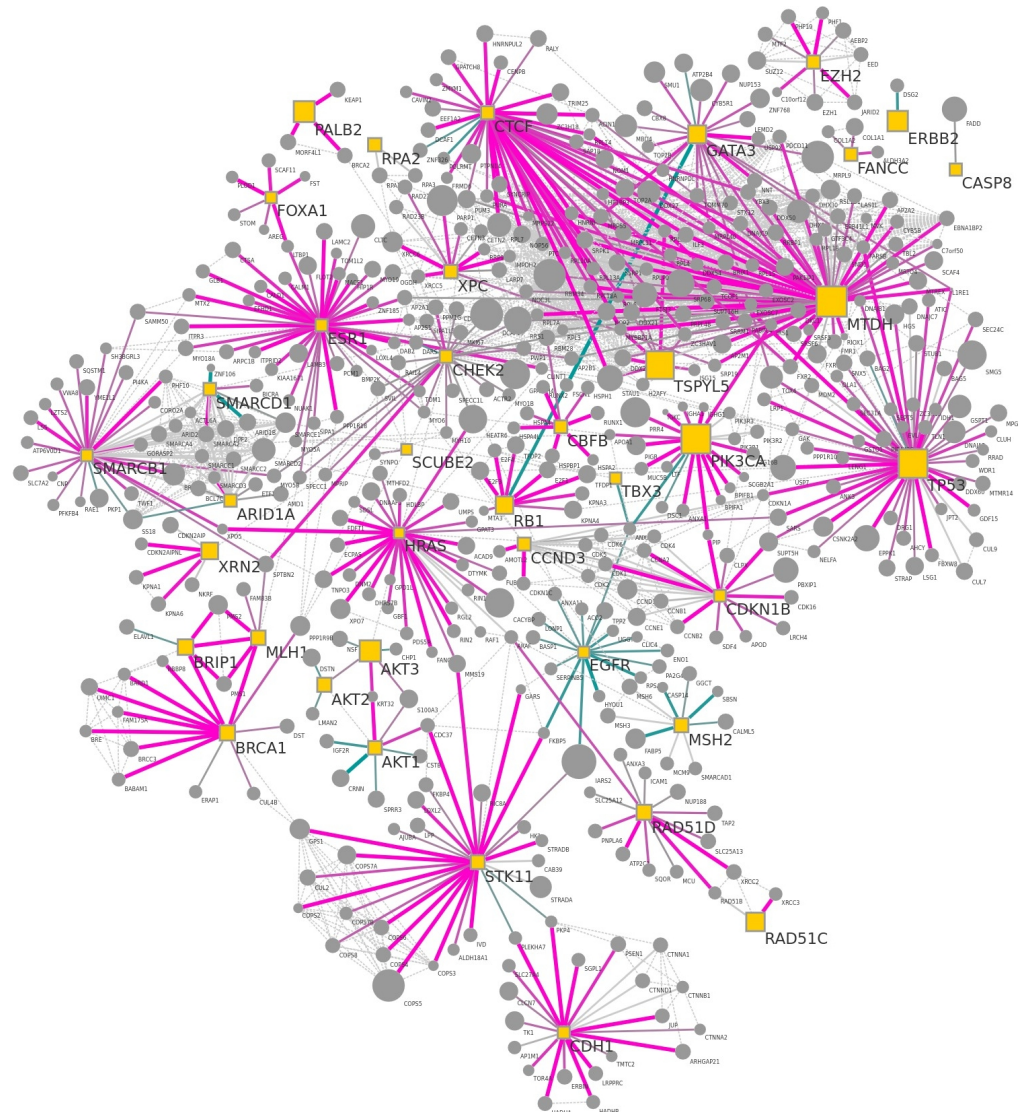

Fig 1. Breast cancer protein-protein interaction network used as example network on NDExEdit. It shows the interactome of the union of all high-confidence PPIs detected across breast cell lines MCF7, MDA-MB-231, and MCF10A. This network is available on NDEx by the UUID: e89ad762-ab4b-11ea-aaef-0ac135e8bacf.

Attribute mappings

A key feature within the data-dependent visualization in Cytoscape is the so-called "attribute-to-visual-mappings" where the values of an attribute are processed by a specific function to generate a new value for the visual representation. Thereby one attribute (or property in the CX context) can be mapped to several visual properties. Cytoscape and the CX-file format distinguish between three kinds of mapping types that can be applied to nodes as well as edges: discrete, continuous, and pass-through.

The values of a property can vary in its data type, which limits the types of mappings that can be applied. For example, for string values, it is not possible to apply a continuous mapping, since by its nature only discrete manifestations are given without any order.

On the other hand, the visual properties vary by type of the value to which they are mapped:

- numerical values for example "NODE_SIZE" or "EDGE_WIDTH"

- colors (in hexadecimal format) for example "NODE_FILL_COLOR"

- string values as in "ELLIPSE" for a "NODE_SHAPE"

- font declarations, including font-family, -style, and -size, are used for example for "NODE_LABEL_FONT_FACE"

NDExEdit limits the choices to select for visual mapping properties to only the applicable types to assure, that only valid mappings can be created. Custom selection tools for colors and fonts are included as well to facilitate the creation of new mappings.

The mappings themselves are stored within the network in the "cyVisualProperties" aspect. This ensures a consistent visual representation of the network on all three platforms, namely NDExEdit, Cytoscape, and NDEx. Furthermore, the modification of the mappings can be continued on either NDEx or Cytoscape.

Discrete mapping

Discrete mappings are the most straightforward type of mappings: to one discrete value of a property, a corresponding mapping value is explicitly assigned. This way, all manifestations of the property can be set individually, but also left blank if no or a default value should be used. Fig 2 shows the discrete mappings of the provided sample network for the properties "Bait" and "BaitBoolean". It shows that each property has only one possible value with already several mappings to visual properties of different data types.

The mapping for the "Bait" property is shown in editing mode with an additional visual property already added using the green plus symbol next to it. The missing mapping value can easily be added using the gray plus symbol or removed with the red "X" button. Also, the visual properties can be removed or restored to the initial value before editing via the provided buttons.

The applied changes can be tested by temporarily showing their effects in the graph by using the magic wand button. All made adjustments can be omitted through the red "X" at the bottom, which leads back to the network overview. Only by actively accepting the changes the new mapping is applied and saved for export.

Continuous mapping

Defining a discrete mapping for continuous values would be tedious since for every value occurring in the attribute a corresponding value for the visual property would be
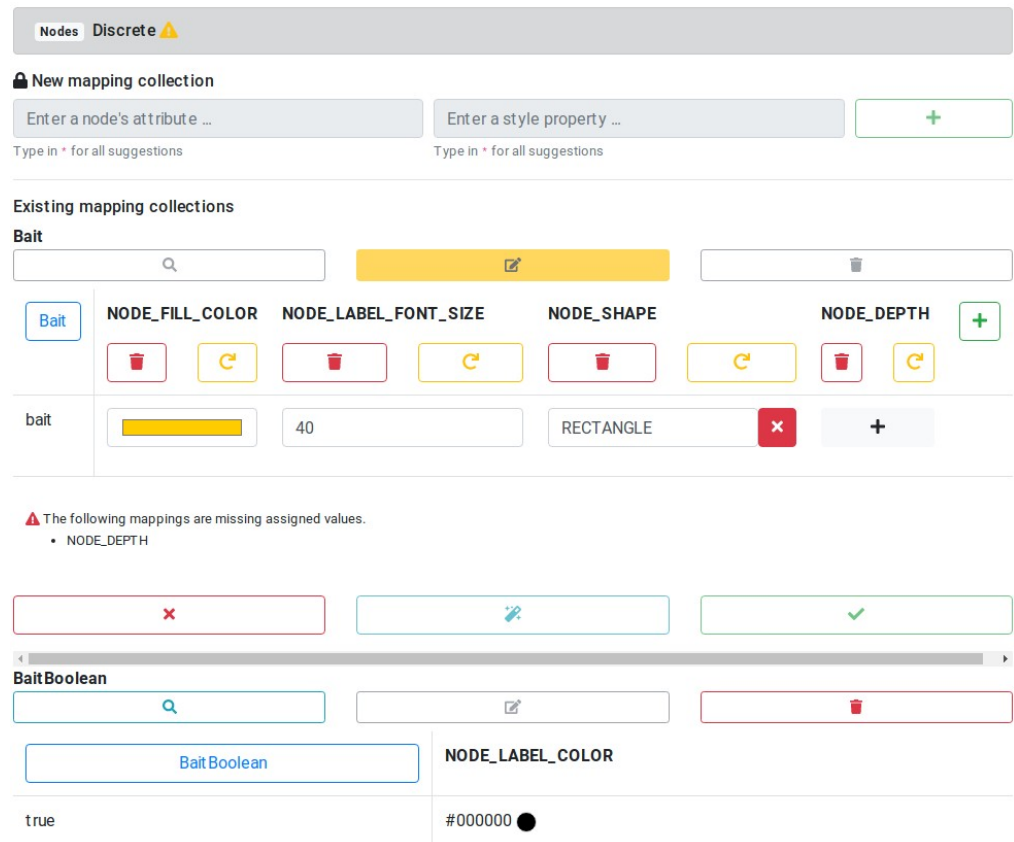
Fig 2. Discrete mapping for node properties. New discrete mappings can be created, existing mappings are shown for the "Bait" property of several visual properties. This includes mappings to colors, numerical and concrete string values.

needed. Continuous mappings relieve one from this burden by defining a function on   140
which basis the values for the visual properties are generated. This function is simply   141
characterized by thresholds for the attribute values with corresponding values for the   142
visual property. All values between two thresholds are then mapped linearly in-between.   143

Continuous mappings can be defined in NDExEdit similarly as discrete mappings,   144
only that the thresholds have to be defined first. Fig 3 shows the continuous mapping of   145
the "diff_score" attribute to two visual properties of the edges. Although several   146
thresholds are defined, the visual properties may not specify all for every visual   147
property. New mapped values can be added using a gray plus button, which appears on   148
moving the cursor over a blank field. Existing ones can not only be deleted but also   149
moved within the visual property by the double-sided arrow next to it.   150

New thresholds can be added with the green plus button at the bottom. This will   151
lead to the new value being attached at the end of the list, therefore the thresholds can   152
be sorted by value. The single thresholds, and corresponding mapping values, can be   153
deleted by the trash bin button next to it. The addition and removal of the visual   154
properties work as for discrete mappings.   155

To facilitate the definition of continuous mappings for an attribute, a histogram of   156
the contained data is displayed along with the editing form, as shown in Fig 4A) for the   157
"diff_score" attribute. It can be seen, that the values lie in the range of -1 and +1. The   158
bin size can be adjusted to get a better overview of the data. This histogram is also   159
shown when the creation of the mapping is finished. Additionally, the different visual   160
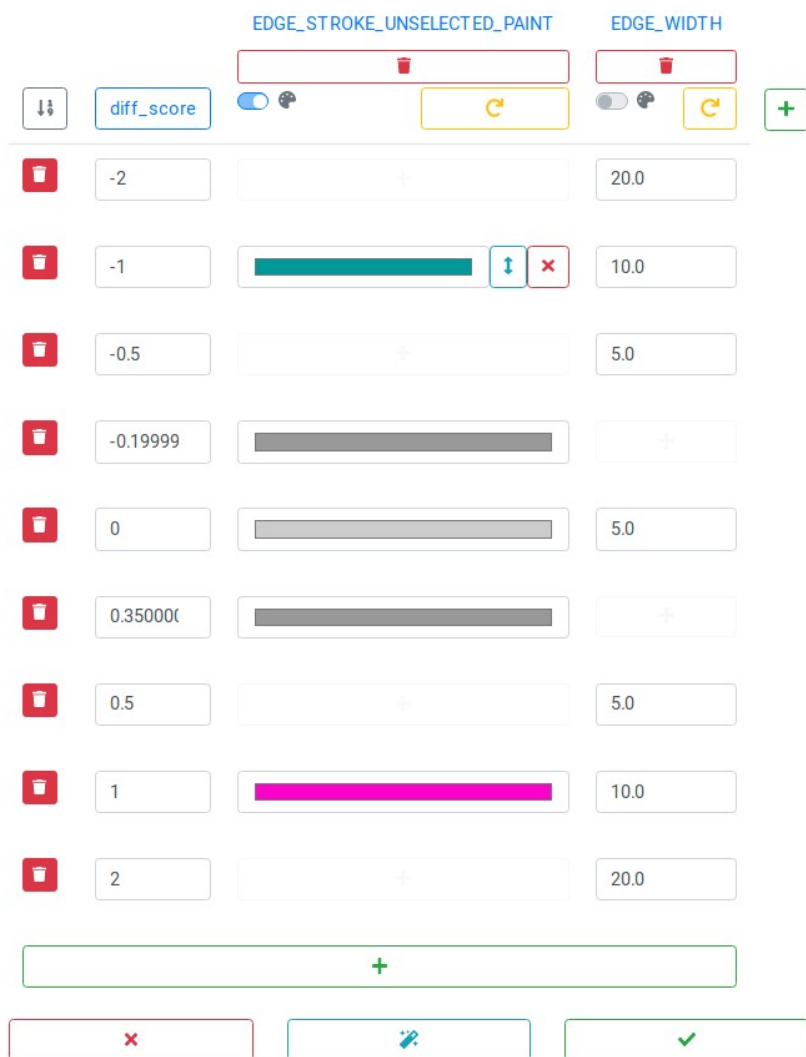
Fig 3. Edit continuous mappings. The score values of the edges are mapped using boundary values, to which colors and numeric values can be assigned. Mapping properties and boundaries can be deleted or new ones added.

properties can be selected to display the resulting mapping. For mappings to colors this shows the corresponding color gradient with marked thresholds (Fig 4B), while for numerical values a graph of the mapping function is displayed (Fig 4C).

Pass-through mapping

Pass-through mappings, as the name suggests, only pass the values of a property through to the mapping attribute. A relatable example is the labels of nodes that are displayed along. Although this mapping could be used to set other mapping properties, such as the node size, this way, in most cases it would be more appropriate to create a continuous mapping, which grants more flexibility afterward.
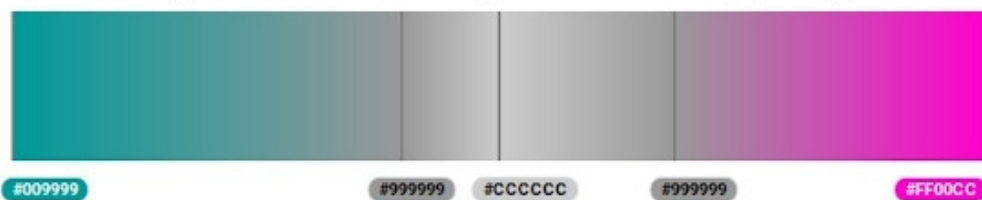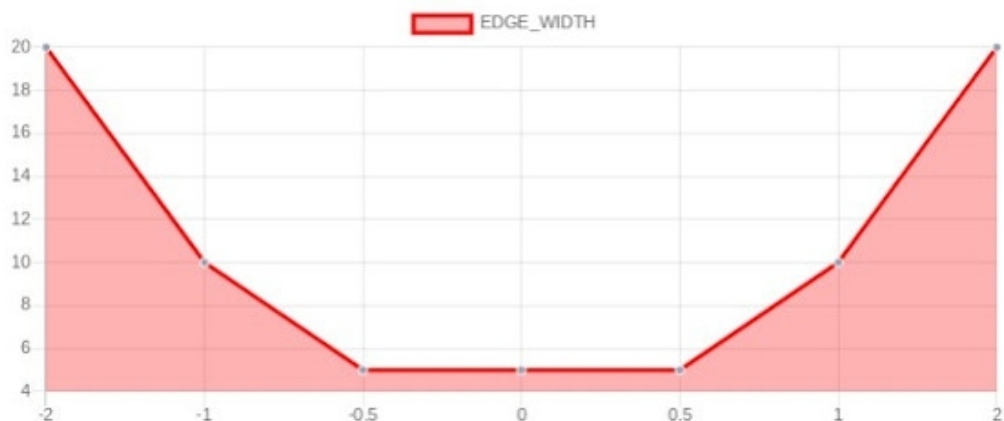
Fig 4. Continuous mapping of values of the edge attribute "diff_score". A) Histogram for the "diff_score" attribute values. B) Continuous mapping of the values to a color gradient with marked boundary values. C) Mapping graph for "diff_score" values to edge width.

Default properties

Mappings can only be created based on the data, which limits the visual representation of the network to the available data. Furthermore, general visual features need to be defined, like the background color of the network. For nodes, edges, and networks those properties can be set there, and then are consequently used as default values to decorate

the networks. They also serve as a fallback when nodes and edges are not covered by the data used for the mappings. 175 176

Graph layout 177

Cytoscape saves the coordinates of the nodes within the network in a dedicated aspect. However, this aspect is only optional, and even not all nodes must have coordinates provided. NDExEdit provides a variety of layout algorithms (Fig 5) to apply to a network, each with a special focus on the networks: 178 179 180 181

- **random**: nodes are distributed randomly across the viewport which enables to roughly explore the network and its content 182 183

- **grid**: nodes are arranged in a grid sorted by the node ids, which puts focus on the nodes 184 185

- **circular**: nodes are arranged in a circle so that the focus lies on the edges between the nodes 186 187

- **concentric**: nodes are arranged in concentric circles which is a more dense representation than the circular layout 188 189

- **hierarchical**: breadth-first arrangement of the network illustrates the topology of the network 190 191

- **force-driven**: cose (Compound Spring Embedder) layout [15] uses a physics simulation to determine node distances and produces a more dense representation of the network topology 192 193 194

- **preset**: initial layout saved within the network allows its restoration 195

The final network layout, including manual refinements, is saved in the cartesian-layout aspect of the exported CX file and is therefore available for the subsequent usage of the network. 196 197 198
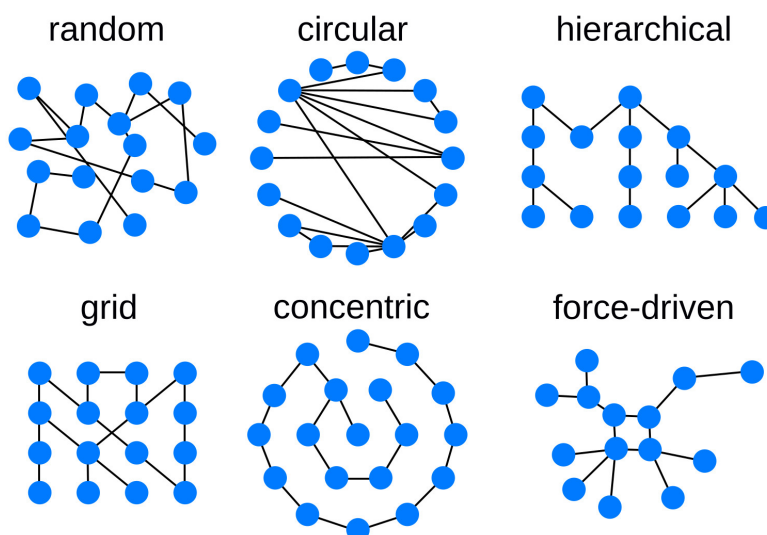

Fig 5. Available graph layout options. Different layout algorithms determine the position of each node, and therefore the overall representation of the network.

Export

NDExEdit provides an option to export the modified networks, including their layout, and visual properties, and mappings as a compatible CX file. Additionally, images in standard formats like PNG and JPEG can be created including a scaling factor to produce more detailed versions than a simple screen capture would allow. Also, the exported image can be set to only capture the viewport, or limited in its dimensions. For images in PNG format, it is also possible to change the background color or leave it transparent.

Limitations and further improvements

Like many browser-based applications, also NDExEdit is limited by the Cross-origin resource sharing (CORS) [16] policy. Since NDExEdit is not hosted at the same domain as the NDEx platform, a direct re-upload of the networks is not possible. This could be solved by hosting a dedicated public backend server simply forwarding the request, but this would impair the privacy of the network data. Alternatively, private instances of NDExEdit and the NDEx platform can be set up, but require the knowledge to do so.

While inspecting several public networks missing features for mappings in general appeared: currently there is no elegant way of defining a mapping, that changes the color at a threshold (Fig 6). Currently, networks resemble this feature by defining a continuous mapping with two close, or even identical values as thresholds. The latter implicates further issues in the validation of the mapping.

Fig 6. Threshold mapping. A rapid change of the color at the threshold is only hardly possible with the current mapping types.

On NDExEdit the specified mappings apply to the whole network, while it would be useful to restrict the mapping to certain sub-networks. Consequently, different mappings could be defined in general and switched on demand by the user. In the CX-format, as well as Cytoscape there already exists a possibility to manage different mappings for sub-networks and views. However, adaption on NDExEdit would require drastic adjustments to the used library for mapping the CX-format to Cytoscape.js.

Taking the idea of managing different mappings even further, would be the possibility to import existing mappings from other networks. This is possible in general, simply by manually editing the CX file and switching the "cyVisualProperties" aspect, but to be able to do it within NDExEdit would further improve the application. This also can be extended to an option to apply predefined visualization templates, such as SBGN [17], STRING [18], or Reactome [19, 20] layouts to a network.

While NDExEdit is intended to be a web application to easily change the visualization of the network dependent on the data, occasionally it would be beneficial to create additional data. For example, if the node degree is not provided as a property, it must be created with other tools to be available for mappings.

Conclusion

Even though NDExEdit only allows visual adjustments to a network, the application illustrates the great potential of web-based solutions for biological research: Users with any operating system can work with NDExEdit without a requirement for installation or account. Since the installation of desktop clients is often restricted due to security

concerns, web-based applications can close this gap and provide access through mobile devices. However, changes to a network's data and algorithmic applications still require a powerful desktop client like Cytoscape.

NDExEdit provides a lightweight interface to explore the contents of networks and facilitates the quick defining of custom visualizations dependent on the data. Networks can be layouted using a variety of built-in algorithms, and refined manually. With compliance to the Cytoscape Exchange format, NDExEdit narrows the gap between desktop clients to create, edit and decorate a network, and platforms to distribute them.

Availability

A live demo is hosted on GitHub Pages at https://frankkramer-lab.github.io/NDExEdit and the corresponding source code for deploying own instances is provided at https://github.com/frankkramer-lab/NDExEdit.

Funding

References

1. Califano A, Butte AJ, Friend S, Ideker T, Schadt E. Leveraging models of cell regulation and GWAS data in integrative network-based association studies. Nature Genetics. 2012;44(8):841–847. doi:10.1038/ng.2355.

2. Chung SS, Pandini A, Annibale A, Coolen ACC, Thomas NSB, Fraternali F. Bridging topological and functional information in protein interaction networks by short loops profiling. Scientific Reports. 2015;5:8540. doi:10.1038/srep08540.

3. Oulas A, Minadakis G, Zachariou M, Sokratous K, Bourdakou MM, Spyrou GM. Systems Bioinformatics: increasing precision of computational diagnostics and therapeutics through network-based approaches. Briefings in Bioinformatics. 2017;doi:10.1093/bib/bbx151.

4. Gligorijević V, Pržulj N. Methods for biological data integration: perspectives and challenges. Journal of The Royal Society Interface. 2015;12(112):20150571. doi:10.1098/rsif.2015.0571.

5. Pavlopoulos GA, Malliarakis D, Papanikolaou N, Theodosiou T, Enright AJ, Iliopoulos I. Visualizing genome and systems biology: technologies, tools, implementation techniques and trends, past, present and future. GigaScience. 2015;4(1):1–27. doi:10.1186/s13742-015-0077-2.

6. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research. 2003;13(11):2498–2504. doi:10.1101/gr.1239303.

7. Pratt D, Chen J, Welker D, Rivas R, Pillich R, Rynkov V, et al. NDEx, the Network Data Exchange. Cell Systems. 2015;1(4):302–305. doi:10.1016/j.cels.2015.10.001.

8. Pillich RT, Chen J, Rynkov V, Welker D, Pratt D. NDEx: A Community Resource for Sharing and Publishing of Biological Networks. Methods in Molecular Biology (Clifton, NJ). 2017;1558:271–301. doi:10.1007/978-1-4939-6783-4_13.

9. Consortium TC. NDEx Network Data Model; 2021. Available from: http://www.home.ndexbio.org/data-model/.

10. Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD. Cytoscape.js: a graph theory library for visualisation and analysis. Bioinformatics. 2016;32(2):309–311. doi:10.1093/bioinformatics/btv557.

11. Jain N, Mangal P, Mehta D. AngularJS: A Modern MVC Framework in JavaScript. Journal of Global Research in Computer Science. 2014;.

12. Bierman G, Abadi M, Torgersen M. Understanding typescript. In: European Conference on Object-Oriented Programming. Springer; 2014. p. 257–281.

13. Otto M. Bootstrap from Twitter; 2021. Developer Blog. Available from: https://blog.twitter.com/developer/en_us/a/2011/bootstrap-twitter.

14. Kim M, Park J, Bouhaddou M, Kim K, Rojc A, Modak M, et al. A protein interaction landscape of breast cancer. Science. 2021;374(6563):eabf3066. doi:10.1126/science.abf3066.

15. Dogrusoz U, Giral E, Cetintas A, Civril A, Demir E. A layout algorithm for undirected compound graphs. Information Sciences: an International Journal. 2009;179(7):980–994. doi:10.1016/j.ins.2008.11.017.

16. Mozilla, individual contributors. Cross-Origin Resource Sharing (CORS).; 2021. Available from: https://developer.mozilla.org/de/docs/Web/HTTP/CORS.

17. Novère NL, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, et al. The Systems Biology Graphical Notation. Nature Biotechnology. 2009;27(8):735–741. doi:10.1038/nbt.1558.

18. Snel B, Lehmann G, Bork P, Huynen MA. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Research. 2000;28(18):3442–3444. doi:10.1093/nar/28.18.3442.

19. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. Nucleic Acids Research. 2020;48(D1):D498–D503. doi:10.1093/nar/gkz1031.

20. Sidiropoulos K, Viteri G, Sevilla C, Jupe S, Webber M, Orlic-Milacic M, et al. Reactome enhanced pathway visualization. Bioinformatics (Oxford, England). 2017;33(21):3461–3467. doi:10.1093/bioinformatics/btx441.