Pre-Print

RCX – an R package adapting the Cytoscape Exchange format for biological networks

Florian Auer[1,*] and Frank Kramer[1]

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

*To whom correspondence should be addressed.

Abstract

Motivation: The Cytoscape Exchange (CX) format is a JSON-based data structure designed for the transmission of biological networks using standard web technologies. It was developed by the network data exchange (NDEx), which itself serves as online commons to share and collaborate on biological networks. The Cytoscape software for the analysis and visualization of biological networks contributes further elements to capture the visual layout within the CX format. However, there is a fundamental difference between web standards and R of how data has to be structured.

Results: Here we present a software package to create, handle, validate, visualize and convert networks in CX format to standard data types and objects within R. Networks in this format can serve as a source for biological knowledge, and also capture the results of the analysis of those while preserving the visual layout across all platforms. The RCX package connects the R environment for statistical computing with platforms for collaboration, analysis and visualization of biological networks.

Availability: RCX is a free and open-source R package, available via GitHub (https://github.com/frankkramer-lab/RCX) and submitted to Bioconductor.

Contact: florian.auer@informatik.uni-augsburg.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Biological networks are a common and widely used resource to capture associations between any types of biological entities such as genes, transcripts, proteins, metabolites, ligands, diseases, or drugs. They often serve as a valuable resource for further investigations and also constitute the result of those. Furthermore, the data formats used for encoding the network information differ heavily depending on the contained data and their intended use.

A variety of public databases provide their, often manually curated, biological knowledge in domain-specific exchange formats. Contrary to the network data used as a source for an analysis the results often can not be stored in a specialized format, especially if in turn it comprises networks. Furthermore, the visualization of resulting networks is not seen as part of the network and omitted.

The Network Data Exchange (NDEx) is an online commons for biological networks (Pratt et al. 2015), that developed the Cytoscape exchange (CX) format. The idea of the data structure is to divide a network into separated aspects, including its visual representation. The data structure for storing those layouts is derived from Cytoscape (Shannon et al. 2003), one of the most popular open-source software tools for the analysis and visualization of biomedical networks. Both, Cytoscape and NDEx use the CX format as a flexible data structure for the exchange of the networks between their platforms.

Users of the statistical programming language R (R Development Core Team 2008) can use existing packages like rBiopaxParser (Kramer et al. 2013) to retrieve biological knowledge from public databases and conduct further analyses. The ndexr package (Auer et al. 2018) interfaces with the NDEx platform to store subsequent results. However, the included data model was missing important features for dealing with biological networks.

2 Features

The RCX package provides a collection of functions to integrate biological networks in CX format into analysis workflows within R. Thereby this package adapts the aspect-oriented design in its data model, which consists of several aspects and sub-aspects, and corresponding properties, that are linked by internal IDs. Additional functions are provided to validate data type usage, aspect structure, and references between the single aspects.

The RCX package not only provides accessibility of networks in CX format, but it also provides conversion to and from objects of iGraph (Csardi and Nepusz 2006) and Bioconductor graph (Gentleman et al. 2021), both widely used libraries for graph manipulation and network analysis (Fig. 1). Furthermore, custom functions for the creation and modification of aspects and RCX objects are all designed to ensure the validity of the networks. This also includes that the meta-data of a network, i.e. an overview of used aspects and its number of elements and IDs, is automatically created and updated on modifications.

The R-based visualization of the networks is congruent with its representation in both, the NDEx platform and Cytoscape. It also can be exported as an HTML file for further use. Since the visual representation is saved as an aspect within the network, it can easily be reused to layout additional networks in the same style without modification.

A key feature of the aspect-oriented design of the CX format is to allow the definition of custom aspects. This package, therefore, was designed concerning easy extensibility, and especially the seamless integration of functions for creation, modification, and conversion of novel aspects.

Detailed documentation and examples can be found in the package manual and vignettes. The Supplementary Materials contain code examples for working with RCX networks and their creation.

3 Implementation

The RCX package is based on several R packages for data processing and graph representation. The CX network files, as well as the HTML visualizations, are read and written with the readr package. The obtained JSON is de- and encoded, and further processed using the jsonlite (Ooms 2014) and tidyr packages (Wickham 2011).

The visualization was realized with the JavaScript library cytoscape.js (Franz et al. 2016) and a custom script to map the visual properties between the in CX used Cytoscape properties to cytoscape.js compatible layout definitions.

4 Conclusion

The RCX package is a freely available R software tool that enables the lossless conversion between the object-oriented JSON format of the CX data structure and the table-like thinking of data in R. The data model was designed to enhance usability and enrich functionality by a better adjustment to fundamental R data structures and adding high-level functions for data manipulation.

Integrated conversion to igraph and Bioconductor compatible graph objects fosters the accessibility to advances network analysis tools. Furthermore, extensibility was increased by facilitating the creation of custom aspects, that cover specialized extensions to the CX data model.

By implementing this software, we ease the task of handling network data available via NDEx within the R Framework for Statistical Computing. Enriched networks as results of investigations and their visualizations can be easily created and translated to the CX format, that connects analysis, visualization, and collaboration.

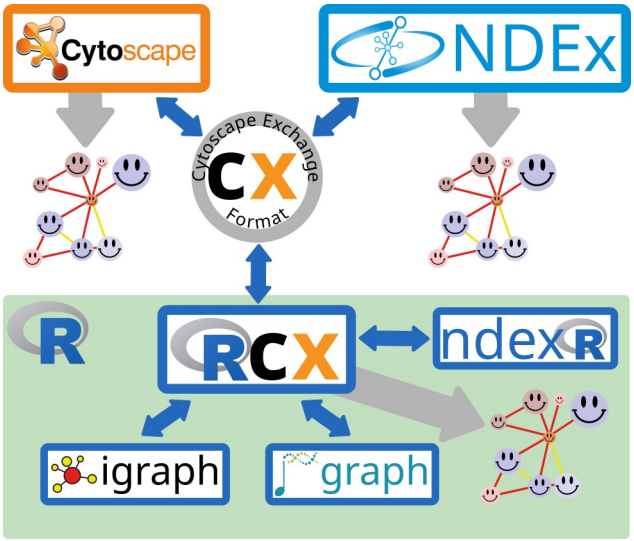

Fig. 1: The RCX package connects the CX transmission format with established analysis libraries in R. The CX format is shared between RCX, Cytoscape and NDEx, while the consistent visual representation remains ensured.

Funding

This work is a part of the Multipath project funded by the German Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) grant FKZ01ZX1508.

Conflict of Interest: none declared.

References

Auer, Florian, Zaynab Hammoud, Alexandr Ishkin, Dexter Pratt, Trey Ideker, Frank Kramer, and Janet Kelso. 2018. "Ndexr-an R Package to Interface with the Network Data Exchange." Bioinformatics (Oxford, England) 34 (4): 716–17. https://doi.org/10.1093/bioinformatics/btx683.

Csardi, G, and T Nepusz. 2006. "The Igraph Software Package for Complex Network Research." InterJournal Complex Systems: 1695.

Franz, Max, Christian T. Lopes, Gerardo Huck, Yue Dong, Onur Sumer, and Gary D. Bader. 2016. "Cytoscape.Js: A Graph Theory Library for Visualisation and Analysis." Bioinformatics 32 (2): 309–11. https://doi.org/10.1093/bioinformatics/btv557.

Gentleman, R, E Whalen, W Huber, and S Falcon. 2021. Graph: A Package to Handle Graph Data Structures. R Package Version 1.70.0.

Kramer, F., M. Bayerlova, F. Klemm, A. Bleckmann, and T. Beissbarth. 2013. "RBiopaxParser--an R Package to Parse, Modify and Visualize BioPAX Data." Bioinformatics 29 (4): 520–22. https://doi.org/10.1093/bioinformatics/bts710.

Ooms, Jeroen. 2014. "The Jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects." ArXiv:1403.2805 [Cs, Stat], March. http://arxiv.org/abs/1403.2805.

Pratt, Dexter, Jing Chen, David Welker, Ricardo Rivas, Rudolf Pillich, Vladimir Rynkov, Keiichiro Ono, et al. 2015. "NDEx, the Network Data Exchange." Cell Systems 1 (4): 302–5. https://doi.org/10.1016/j.cels.2015.10.001.

R Development Core Team. 2008. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. "Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks." Genome Research 13 (11): 2498–2504. https://doi.org/10.1101/gr.1239303.

Wickham, Hadley. 2011. "The Split-Apply-Combine Strategy for Data Analysis." Journal of Statistical Software 40 (April): 1–29. https://doi.org/10.18637/jss.v040.i01.