The Journal of Open Source Software

Building, Importing, and Exporting GEXF Graph Files with rgexf

George G Vega Yon[1]

1 Department of Preventive Medicine, University of Southern California

DOI: 10.21105/joss.03456

Software

- Review
- Repository
- Archive

Editor: Fabian Scheipl

Reviewers:

- @jonjoncardoso
- @corneliusfritz

Submitted: 13 June 2021
Published: 07 July 2021

Summary

First introduced in 2012, the **rgexf** package for the R programming language was the first effort to make the Graph Exchange XML Format (GEXF) (Heymann et al., 2009) specification available to the **R** world. With more than 500,000 downloads[1], it is one of the most popular ways to incorporate GEXF files into the R programming language environment.

Developed by the Gephi Core Group (Bastian et al., 2009), the GEXF specification is a flexible and widely used format to describe graphs. Although it has not been updated since 2009, the GEXF format has been introduced to several tools and programming environments. A few examples include:

- The python library **networkx** (Hagberg et al., 2008)
- The stand-alone software **Cytoscape** (Smoot et al., 2010)
- The JavaScript library **sigma.js** https://simga.js
- The java library **gexf4j** https://github.com/francesco-ficarola/gexf4j)
- The JavaScript library **gexf-js** https://github.com/raphv/gexf-js

Besides the **rgexf** package, other R packages provide functions that interact with GEXF files:

- **sigmajs**: Interface to 'Sigma.js' Graph Visualization Library (Coene, 2018)
- **vkR**: Access to VK API via R (Sorokin, 2020)
- **microeco**: Microbial Community Ecology Data Analysis (Liu et al., 2021)
- **netCoin**: Interactive Analytic Networks (Escobar & Martinez-Uribe, 2020)

Nevertheless, the **rgexf** package continues to be the de-facto tool to interact with GEXF files in **R**.

Statement of Need

This R package has been serving the scientific community for many years now. Scientists and data analysts across the board have been using **rgexf** to enhance their analyses by smoothly moving between **R** and other applications used for graph visualization. Some concrete examples include gene networks (Kauffman et al., 2018; Starr et al., 2017), interactions among species (Leclerc et al., 2018), and social networks (Alsaedi et al., 2016).

[1] According to the https://cranlogs.r-pkg.org/ website, as of June 14, 2021.

Features

Beyond reading and writing GEXF files from within R, the **rgexf** R package has various other features that can help to create beautiful network visualizations, in particular:

- Using gexf objects–the main class implemented in rgexf–users can create GEXF objects from scratch, adding and removing nodes and edges–including features–as needed.
- Users of the **igraph** package can directly convert objects between `gexf` and `igraph` classes.
- Thanks to the **gexf-js** javascript library, users can immediately visualize their network objects in the web browser.

Because of these and other reasons, the **rgexf** package has been featured in many scientific papers, stating the great utility that this R package has provided to the community. The **rgexf** package is available in the Comprehensive R Archive Network (CRAN) and the project repository at https://github.com/gvegayon/rgexf.

References

Alsaedi, N., Burnap, P., & Rana, O. (2016). Temporal TF-IDF: A High Performance Approach for Event Summarization in Twitter. *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 515–521. https://doi.org/10.1109/WI.2016.0087

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media*, *3*.

Coene, J.-P. (2018). Sigmajs: An r htmlwidget interface to the sigma.js visualization library. *Journal of Open Source Software*, *3*(28), 814. https://doi.org/10.21105/joss.00814

Escobar, M., & Martinez-Uribe, L. (2020). Network coincidence analysis: The netCoin R package. *Journal of Statistical Software*, *93*(11), 1–32. https://doi.org/10.18637/jss.v093.i11

Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using NetworkX*. Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Heymann, S., Bastian, M., Jacomy, M., Maussang, C., Rohmer, A., Bilcke, J., & Jacomy, A. (2009). GEXF file format. *GEXF Working Group,[online] C2009 [cit. 2013-05-07]. Available at WWW:< Http://Gexf. Net/Format*.

Kauffman, K. M., Hussain, F. A., Yang, J., Arevalo, P., Brown, J. M., Chang, W. K., VanInsberghe, D., Elsherbini, J., Sharma, R. S., Cutler, M. B., Kelly, L., & Polz, M. F. (2018). A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. *Nature*, *554*(7690), 118–122. https://doi.org/10.1038/nature25474

Leclerc, C., Courchamp, F., & Bellard, C. (2018). Insular threat associations within taxa worldwide. *Scientific Reports*, *8*(1), 6393. https://doi.org/10.1038/s41598-018-24733-0

Liu, C., Cui, Y., Li, X., & Yao*, M. (2021). Microeco: An r package for data mining in microbial community ecology. *FEMS Microbiology Ecology*, *97*(2), fiaa255. https://doi.org/10.1093/femsec/fiaa255

Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., & Ideker, T. (2010). Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, *27*(3), 431–432. https://doi.org/10.1093/bioinformatics/btq675

Sorokin, D. (2020). *vkR: Access to VK API via r*. https://CRAN.R-project.org/package=vkR

Starr, T. N., Picton, L. K., & Thornton, J. W. (2017). Alternative evolutionary histories in the sequence space of an ancient protein. *Nature*, *549*(7672), 409–413. https://doi.org/10.1038/nature23902