OXFORD

Databases and ontologies

ndexr—an R package to interface with the network data exchange

Florian Auer[1,*], Zaynab Hammoud[1], Alexandr Ishkin[2], Dexter Pratt[3], Trey Ideker[3,4] and Frank Kramer[1]

[1]Department of Medical Statistics, University Medical Center Göttingen, Göttingen 37099, Germany, [2]Discovery Science, Clarivate Analytics, Boston, MA 02210, USA, [3]Department of Medicine, University of California San Diego, La Jolla, CA 92093, USA and [4]Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA

*To whom correspondence should be addressed.

Associate Editor: Janet Kelso

Abstract

Motivation: Seamless exchange of biological network data enables bioinformatic algorithms to integrate networks as prior knowledge input as well as to document resulting network output. However, the interoperability between pathway databases and various methods and platforms for analysis is currently lacking. The Network Data Exchange (NDEx) is an open-source data commons that facilitates the user-centered sharing and publication of networks of many types and formats.

Results: Here, we present a software package that allows users to programmatically connect to and interface with NDEx servers from within R. The network repository can be searched and networks can be retrieved and converted into igraph-compatible objects. These networks can be modified and extended within R and uploaded back to the NDEx servers.

Availability and implementation: ndexr is a free and open-source R package, available via GitHub (https://github.com/frankkramer-lab/ndexr) and Bioconductor (http://bioconductor.org/packages/ndexr/).

Contact: florian.auer@med.uni-goettingen.de

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

1 Introduction

Networks are a convenient representation of complex relations and interactions and are commonly used in a wide range of fields in biology. The information represented in networks can range from knowledge on molecular interactions and cellular pathways (Hucka *et al.*, 2003; Kohn, 1999) to the results of bioinformatic methods for network reconstruction (Margolin *et al.*, 2006).

Public databases Reactome (Fabregat *et al.*, 2016) provide access to the rapidly increasing body of biological pathway knowledge and are a well-established source for hypothesis generation and testing. Existing packages, like rBiopaxParser (Kramer *et al.*, 2013), already enable users of the statistical programming language R (R Development Core Team, 2008) to work with biological pathway data.

Researchers, however, still face challenges in dealing with network complexity, diverse data formats, the integration of network analysis tools and methods to share and collaborate on pathway data.

2 Network data exchange

The network data exchange (NDEx) is an open-source software framework to manipulate, store and exchange networks of various types and formats (Pratt *et al.*, 2015).

The NDEx can be used to upload, share and distribute network data, facilitating the creation and curation of networks by users and communities. It can serve as both a source for networks consumed by applications and a destination for the networks that they produce.

The server can be accessed by end users via a web interface and by programs using a relational state transfer application programming interface (REST API) (Fielding and Taylor, 2002).

3 Features

This package provides an interface to NDEx installations from within R and enables a seamless transition from data acquisition to statistical analysis.

Using the NDEx REST API, this package provides an interface to the public NDEx server, as well as private installations, enabling programs to upload, download or modify biological networks. The package also provides classes to implement the cytoscape cyberinfrastructure (CX) format, a flexible, modular and extensible data structure for the transmission of networks. Furthermore, it provides conversion to objects of the *iGraph* package (Csardi and Nepusz, 2006), a widely used library for graph manipulation and network analysis.

A typical workflow illustrating the most important features of this package is described in Figure 1 and might include following steps:

Browse, search and query NDEx to find a network of interest, either as an authenticated user or as an anonymous visitor.

Download a network into R and convert it to built-in data structures resembling the CX structure or to an iGraph object.

1. Perform some network analysis (3a) or apply a typical bioinformatics workflow by integrating additional data and subsequently selecting a subnetwork (3b).
2. Upload the newly created network to the NDEx server.
3. Share the preliminary network only with certain people (5a) or groups of persons (5b).
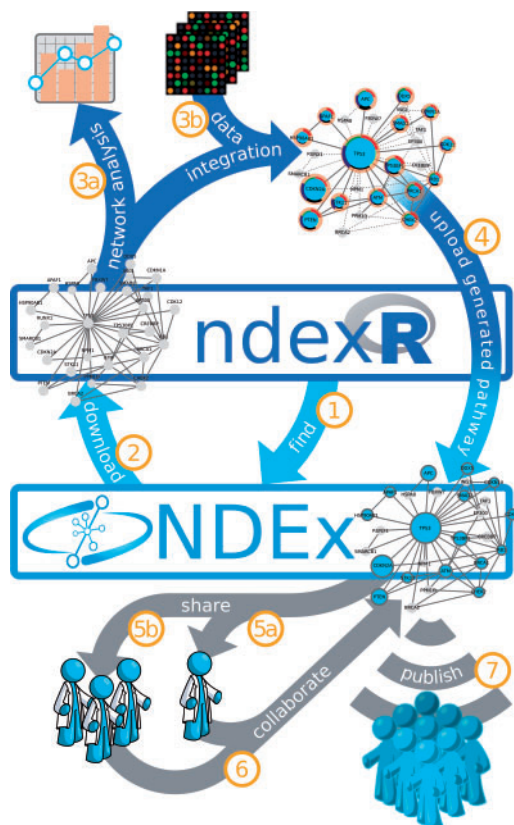
4. Collaborate with other people or groups to amend and complete the network.

Reveal the network to the public and/or provide it as supplement along a publication.

Detailed documentation and examples can be found in the package vignettes and manual. The Supplementary Materials contain code examples for common procedures to interact with the public NDEx server.

4 Implementation

The *ndexr* package is based on several R packages for data processing, internet communication and graph representation. Network connections to the NDEx server APIs are handled using the *httr* package. Network data is de- and encoded, as well as transformed into the different interchangeable data structures using the *jsonlite* package and the *plyr* and *tidyr* packages (Wickham, 2011).

This package also provides classes specifically tailored to cope with NDEx and analysis: The package supports NDEx versions 1.3 and 2, and new features and API specifications will be included in regular package updates. Additionally, it is possible to switch and modify the API configuration manually.

5 Conclusion

The ndexr package is a freely available R software tool which enables users to connect to and interact with NDEx servers. Package methods enable networks to be found via queries, retrieved and be converted into igraph compatible graph objects. These graphs can be used and modified within R and uploaded to the NDEx servers. By implementing this software, we ease the task of retrieving and using network data available via NDEx within the R Framework for Statistical Computing.

Fig. 1. Workflow using the NDEx server and the ndexr package. The co-operation of both software tools enables a faster and more efficient way to acquire, alter, share and publish networks

References

Csardi,G. and Nepusz,T. (2006) The igraph software package for complex network research. *InterJournal Complex Systems*, 1695.

Fabregat,A. *et al.* (2016) The reactome pathway knowledgebase. *Nucleic Acids Res.*, 44, D481–D487.

Fielding,R.T. and Taylor,R.N. (2002) Principled design of the modern web architecture. *ACM Trans. Internet Technol.*, 2, 115–150.

Hucka,M. *et al.* (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19, 524–531.

Kohn,K.W. (1999) Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell*, 10, 2703–2734.

Kramer,F. *et al.* (2013) rBiopaxParser—an R package to parse, modify and visualize BioPAX data. *Bioinformatics*, 29, 520–522.

Margolin,A.A. *et al.* (2006) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics*, 7, S7.

Pratt,D. *et al.* (2015) NDEx, the network data exchange. *Cell Syst.*, 1, 302–305.

R Development Core Team. (2008) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Wickham,H. (2011) The split-apply-combine strategy for data analysis. *J. Stat. Softw.*, 40, 1–29.