BIOINFORMATICS APPLICATIONS NOTE

Vol. 27 no. 1 2011, pages 140–141
doi:10.1093/bioinformatics/btq606

Systems biology

Advance Access publication November 11, 2010

QuACN: an R package for analyzing complex biological networks quantitatively

Laurin A. J. Mueller[1,*], Karl G. Kugler[1], Andreas Dander[2], Armin Graber[1] and Matthias Dehmer[1]

[1]Department of Biomedical Sciences and Engineering, Institute for Bioinformatics and Translational Research, University for Health Sciences, Medical Informatics and Technology (UMIT), Hall in Tirol and [2]Biocenter, Division for Bioinformatics, Innsbruck Medical University, Innsbruck, Austria

Associate Editor: Trey Ideker

ABSTRACT

Motivation: Network-based representations of biological data have become an important way to analyze high-throughput data. To interpret the large amount of data that is produced by different high-throughput technologies, networks offer multifaceted aspects to analyze the data. As networks represent biological relationships within their structure, it turned out to be fruitful to analyze their topology. Therefore, we developed a freely available, open source R-package called Quantitative Analysis of Complex Networks (QuACN) to meet this challenge. QuACN contains different, information-theoretic and non-information-theoretic, topological network descriptors to analyze, classify and compare biological networks.

Availability: QuACN is freely available under LGPL via CRAN (http://cran.r-project.org/web/packages/QuACN/).

Contact: laurin.mueller@umit.at

Received and revised on August 31, 2010; accepted on October 25, 2010

1 INTRODUCTION

To analyze structural data, networks have become more and more a key technology in Systems Biology (Palsson, 2009). Examples of underlying networks are biological pathways, protein–protein interactions, correlation networks, mutual information networks, etc. (Junker and Schreiber, 2008). Clearly, graph theoretical concepts like classical graph descriptors, e.g. degree distribution, clustering coefficients or betweenness, can be used to analyze complex biological networks. These descriptors help to generate statements about network characteristics like hubs, clusters, modules, etc. (Junker and Schreiber, 2008).

Additionally, there exists a large number of more sophisticated and expressive graph measures using information theory. For example, Todeschini *et al.* (2002) lists a large number of descriptors that can be used to analyze molecular networks. They can also be utilized to analyze complex biological networks, e.g. gene networks, protein–protein interaction networks, etc. However, a study to investigate the performance of these descriptors on biological networks does not exist so far. Moreover, the biological meaning of these descriptors (Todeschini *et al.*, 2002), has not been explored until now.

*To whom correspondence should be addressed.

Topological descriptors can be used to classify (Mueller *et al.*, 2010) and compare (Kugler *et al.*, 2010) correlation networks created from microarray data and to characterize chemical compounds in order to identifying potential drug targets (Dehmer *et al.*, 2009a), etc.

We implemented a selection of sophisticated topological network measures and provide them as a package for the free and open source software environment R. R offers several packages and methods to infer networks from biological data [e.g. minet (Meyer *et al.*, 2008) or WGCNA (Langfelder and Horvath, 2010)], which can be directly analyzed with QuACN.

Topological descriptors can be used to classify and compare correlation networks created from micro-array data as published in Mueller *et al.* (2010) and Kugler *et al.* (2010), to characterize chemical compounds identifying potential drug targets (Dehmer *et al.*, 2009a), etc.

Little is known what network characteristics can be captured using topological descriptors exactly. We recommend QuACN to investigate the behavior of topological network measures on large complex networks. Further, we hope that the package will be helpful when exploring questions concerning the structure of biological networks in the context of systems biology.

2 METHODS

The R-package QuACN contains various topological network descriptors, that can be used to analyze biological networks, out of following classes based on Dehmer *et al.* (2009a):

(1) *Descriptors based on distances in a graph:* this class contains measures using distances to describe the networks structure.

(2) *Descriptors based on other graph invariants:* the descriptors in this class use other graph invariants than distances (e.g. degree, number of nodes, number of edges, etc.).

(3) *Partition-based graph entropy descriptors:* these measures use an arbitrary graph invariant and an equivalence criteria to induce partitions. A probability value is calculated for each partition to determine the entropy, based on the entropy formula of Shannon (Shannon and Weaver, 1997).

(4) *Parametric graph entropy measures:* to determine the entropy measures of this class (Dehmer *et al.*, 2009b), by assigning a probability value to each vertex of the network, using the so-called information functionals.

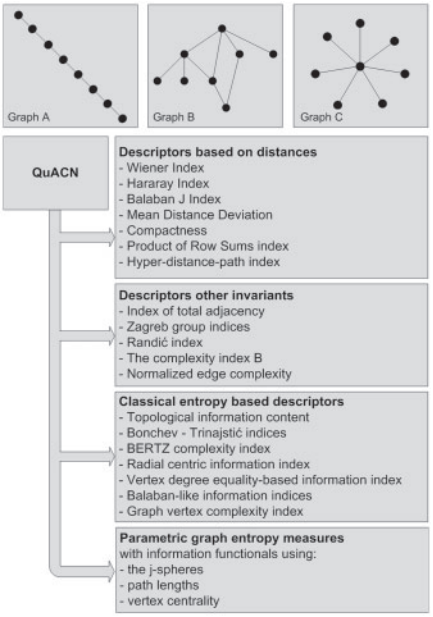

Fig. 1. This figure gives a graphical overview of the topological network descriptors integrated in QuACN. It also shows three small sample graphs with obviously different topology.

Table 1. This table shows a selection of network measures applied to the sample graphs in Figure 1

Network descriptor	Graph A	Graph B	Graph C
Wiener index	84	54	49
Balaban J index	2.53006	2.277706	5.136596
Zagreb index	24	70	49
Bonchev - Trinajstić index	3.529638	1.674978	0.3818656
Balaban-like information index	18.82699	9.928763	17.54814
Dehmer entropy[a]	41.06296	62.55411	68.7402

[a]Parametric entropy measure with an information functional using j-sphere cardinalities (Dehmer *et al.*, 2009b).

An overview of the implemented methods is shown in Figure 1. To demonstrate the performance of QuACN, Table 1 shows some selected network measures applied to small sample graphs with different topological characteristics (see Fig. 1). One can see that the descriptors capture different network topologies (e.g. branching) in a different manner.

We chose to use R because it is open source and there already exist packages to handle graphs [e.g. *graph* (Gentleman *et al.*, 2009) or *igraph* (Csardi and Nepusz, 2006)]. To analyze graphs or networks with QuACN they have to be represented by a *graphNEL*-object of the *Bioconductor*-package *graph*. We picked *graphNEL* as the base object, because this offers several advantages. First, the *graph* package provides a variety of methods to import already existing networks using adjacency matrices, node-edge list or the Graph Markup Language (GML). Secondly, it is possible to convert *graphNEL* objects into *igraph*-objects and vice versa. This offers the possibility to use the wide range of graph-theoretical methods provided by the *igraph* package.

3 DISCUSSION

There exist a few stand-alone applications (Lee *et al.*, 2002; Todeschini *et al.*, 2003) to calculate topological descriptors, which are designed to work with molecular networks. Due to the fact that R is commonly used in the field of bioinformatics and that biological networks are not stored in molecular file formats, a new R package has been developed. R offers a wide range of methods to import existing networks and biological data.

QuACN contains a selection of topological network descriptors to analyze, classify and compare complex networks. We want to direct one's attention to the parametric graph entropy measures that are, at the moment, only available within this package. They facilitate to exploit machine learning techniques to derive hypothesis from complex biological datasets.

QuACN is a a powerful open source package for the analysis of complex biological networks. In future work, we plan to apply the integrated measures on various biological research questions, and extending the range of functions with new promising descriptors for the coming versions of QuACN.

ACKNOWLEDGEMENT

We want to thank Matthias Wieser who supported us during the implementation.

Funding: This work was supported by the Tiroler Wissenschaftsfonds and by the GenAU BIN III project.

Conflict of Interest: none declared.

REFERENCES

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

Dehmer,M. *et al.* (2009a) A large scale analysis of information-theoretic network complexity measures using chemical structures. *PLoS ONE*, **4**, e8057.

Dehmer,M. *et al.* (2009b) On entropy-based molecular descriptors: statistical analysis of real and synthetic chemical structures. *J. Chem. Inf. Model.*, **49**, 1655–1663.

Gentleman,R. *et al.* (2009) *graph: graph: A package to handle graph data structures*. R package version 1.26.0. Available at http://cran.r-project.org/web/packages/graph/.

Junker,B.H. and Schreiber,F. (2008) *Analysis of Biological Networks (Wiley Series in Bioinformatics)*. Wiley-Interscience, Hoboken, New Jersy.

Kugler,K.G. *et al.* (2010) A novel majority vote count algorithm for integrative analysis of association networks. In *Conference on Bioinformatics & Computational Biology (BIOCOMP'10)*. CSREA Press, Las Vegas/USA, **I**, pp. 62–67.

Langfelder,P. and Horvath,S. (2010) *WGCNA: Weighted Gene Co-Expression Network Analysis*. R package version 0.92-3. Available at http://cran.r-project.org/web/packages/WGCNA/.

Lee,S.K. *et al.* (2002) The preadme approach: Web-based program for rapid prediction of physico-chemical, drug absorption and drug-like properties. In *euro QSAR 2002 - Designing Drugs and Crop Protectants: Processes Problems and Solutions*, Blackwell Publishing, Massachusetts, pp. 418–420.

Meyer,P.E. *et al.* (2008) minet: a r/bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics*, **9**, 461.

Mueller,L.A. *et al.* (2010) Network-based approach to classify disease stages of prostate cancer using quantitative network measures. *Conference on Bioinformatics & Computational Biology (BIOCOMP'10)*. CSREA Press, Las Vegas/USA, **I**, pp. 55–61.

Palsson,B. (2009) Metabolic systems biology. *FEBS Lett.*, **583**, 3900–3904.

Shannon,C.E. and Weaver,W. (1997) *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL, USA.

Todeschini,R. *et al.* (2002) *Handbook of Molecular Descriptors*. Wiley-VCH. Weinheim, Germany.

Todeschini,R. *et al.* (2003) *Software Dragon: Calculation of Molecular Descriptors*. Department of Environmental Sciences. Talete, Milan, Italy.