

SOFTWARE

Open Access

Structural Measures for Network Biology Using QuACN

Laurin AJ Mueller¹, Karl G Kugler¹, Armin Graber¹, Frank Emmert-Streib² and Matthias Dehmer^{1*}

Abstract

Background: Structural measures for networks have been extensively developed, but many of them have not yet demonstrated their sustainability. That means, it remains often unclear whether a particular measure is useful and feasible to solve a particular problem in network biology. Exemplarily, the classification of complex biological networks can be named, for which structural measures are used leading to a minimal classification error. Hence, there is a strong need to provide freely available software packages to calculate and demonstrate the appropriate usage of structural graph measures in network biology.

Results: Here, we discuss topological network descriptors that are implemented in the R-package QuACN and demonstrate their behavior and characteristics by applying them to a set of example graphs. Moreover, we show a representative application to illustrate their capabilities for classifying biological networks. In particular, we infer gene regulatory networks from microarray data and classify them by methods provided by QuACN. Note that QuACN is the first freely available software written in R containing a large number of structural graph measures.

Conclusion: The R package QuACN is under ongoing development and we add promising groups of topological network descriptors continuously. The package can be used to answer intriguing research questions in network biology, e.g., classifying biological data or identifying meaningful biological features, by analyzing the topology of biological networks.

Background

Understanding the structure and dynamics of biological systems has been a major task in systems biology [1]. In the early years of computational biology, the main task was to investigate the individual properties of intracellular components and collect this information in large databases [2]. Pálsson defines biological systems as interactions of their components [3]. Furthermore, the development of high-throughput technologies made it possible to study these complex systems in a quantitative manner [4]. Moreover, gene networks, whose nodes represent gene products and the edges correspond to molecular interactions, serve as means to study the biological function by representing and analyzing high-throughput data [2].

Network inference plays a major role in network biology, as there exist various methods to infer networks from high-throughput data [5-9]. By using the WGCNA package [10] it is possible to create correlation networks. One can use the minet package [6] to infer networks based on mutual information. Other packages [11-13] offer methods to infer networks using different kinds of graphical models. Moreover, Altay and Emmert-Streib introduced the C3NET algorithm to infer the conservative causal core of gene networks and compared them to other approaches [5]. Their study shows the importance of correctly creating robust and valid networks from biological data. Note that it is crucial to choose suitable methods for inferring networks from biological data, in order to take the nature and constraints of the underlying problem into account [5]. After inferring gene networks, it is often important to analyze them structurally to conclude statements about the underlying topology [14,15]. Moreover, the structural analysis of biological networks can be useful to extract biological knowledge that may not be revealed

* Correspondence: matthias.dehmer@umit.at

¹Institute for Bioinformatics and Translational Research, Department of Biomedical Sciences and Engineering, University for Health Sciences, Medical Informatics and Technology (UMIT), EWZ 1, Hall in Tirol, Austria
Full list of author information is available at the end of the article

by studying the raw data [16]. Typical problems aim at identifying of topological interesting nodes or characterizing the networks by means of their structure. Therefore, we provide an R package called QuACN [17] providing a selection of new topological network descriptors. Such descriptors are numerical graph invariants that quantitatively characterize the structure of the underlying network. Note, that the authors use the words *descriptor*, *measure*, or *index* as synonym for topological network descriptors.

Quantifying the complexity of networks appears in different scientific disciplines and has been a challenging research topic during the last decades [15]. Importantly, little is known about the *structural interpretation* of topological network descriptors [14,15]. This relates to information-theoretic measures [14,18-21] that had been used to determine the entropy of the graph topology. Other topological network descriptors had been used also in mathematical and medical chemistry including drug design to analyze and characterize the structure of chemical compounds (QSAR/QSPR) [15,22-24].

In more biologically motivated work, Xia et al. [25] used the vertex degree of protein-protein interaction (PPI) networks to correlate the structural complexity of proteins and the organismal complexity with the complexity of the underlying PPI network. They show that the PPI domain coverage significantly correlates with the vertex degrees of the PPI networks [25]. In another study, Mazurie et al. [26] used different network measures to link the structure and complexity of metabolic reactions (interacting pathways) to the phylogeny of species. Their results show that a small set of descriptors reproduces the phylogenetic distances accurately [26].

Numerous network measures have been developed, but it would be out of the scope of this paper to explain them in detail. For further investigation see the recently and up to date review due to Dehmer and Mowshowitz [27]. Apart from information-theoretic measures, Todeschini et al. [24] provides a compelling overview of available network descriptors. But from [24], the feasibility and properties of a large number of descriptors remain untackled.

QuACN provides a selection of topological network descriptors. It offers the possibility to apply the indices in a standardized and intuitive manner. Thus, it can support the scientific community to investigate these methods in different kinds of biological applications. A typical setup for a study to analyze biological networks structurally is illustrated in Figure 1. It shows a general workflow to analyze microarray studies using a network approach with topological network measures.

Of course, there also exist freely available tools, e.g., PowerMV [28] or JOELib [29] to calculate network descriptors. However, these tools are designed for

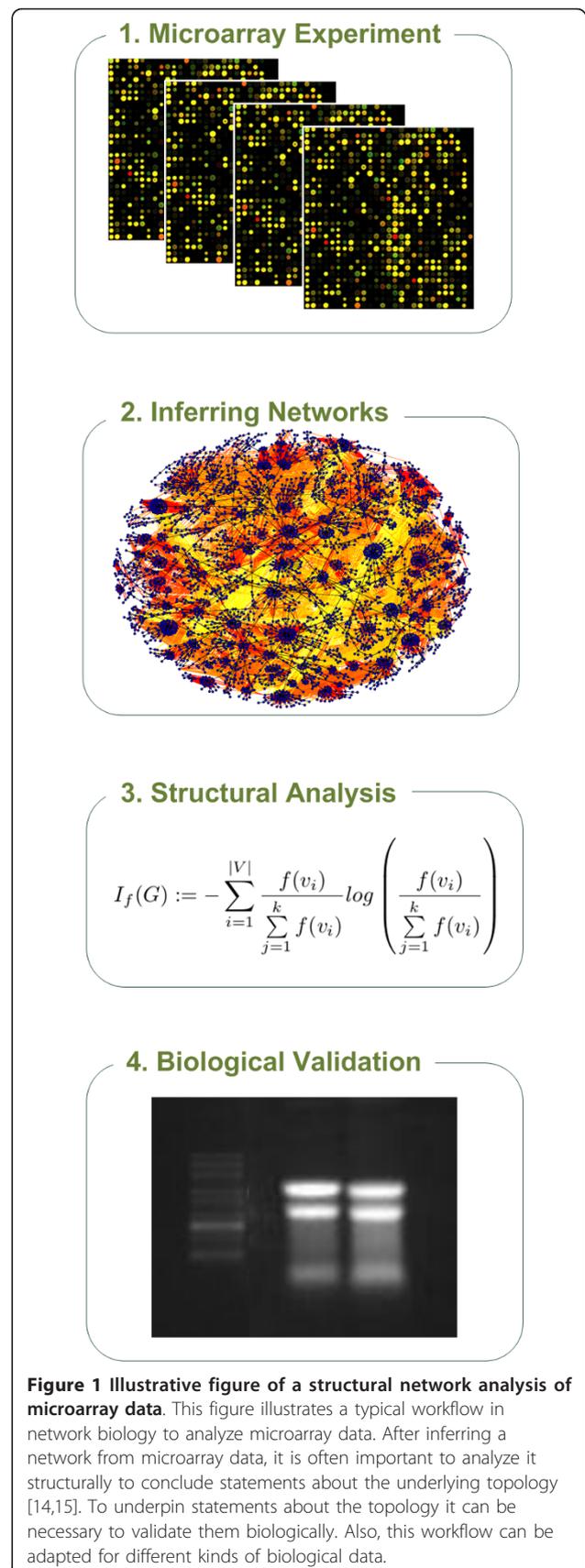


Figure 1 Illustrative figure of a structural network analysis of microarray data. This figure illustrates a typical workflow in network biology to analyze microarray data. After inferring a network from microarray data, it is often important to analyze it structurally to conclude statements about the underlying topology [14,15]. To underpin statements about the topology it can be necessary to validate them biologically. Also, this workflow can be adapted for different kinds of biological data.

quantitative structure-activity relationship (QSAR). Thus, they do not support common exchange standards for biological data. Compared to commercial software tools as Dragon [30] or PreADMET [31], QuACN is published under an open source license (LPGL) and freely available. Therefore, it offers the possibility to adjust and further develop the existing indices or even add additional descriptors to the package. Compared to the R-packages igraph [32] and RBGL [33], which contain a few basic descriptors, QuACN contains a selection of more sophisticated network descriptors (i.e., the group of entropy-based descriptors). To our best knowledge, it is the only available software package that contains sophisticated measures such as the parametric graph entropies (Dehmer entropy) [34]. We recommend QuACN to investigate large-scale complex networks. Further, we expect that the package will be helpful for exploring questions concerning the structure of biological networks in the context of systems biology.

Generally, quantitative network analysis [35] is a non-trivial task, since it is necessary to understand the methods in detail to interpret the results correctly. This manuscript addresses readers who want to analyze networks structurally. Its aim is to guide the reader to correctly apply the methods provided by QuACN [17]. This manuscript does not deal with the issue of inferring robust and valid networks. Neither does it explain the network measures in detail nor how to interpret the results of the topological networks descriptors, as this would go beyond the scope of this paper. Dehmer et al. dealt with these questions extensively [15,27]. This paper is structured as follows: The section *Implementation* gives an overview about the topological network descriptors, implemented in the R-package QuACN. The section *Results and Discussion* illustrates how to apply the topological descriptors to concrete networks. Also, we show the behavior of selected measures using small example graphs. Moreover, we demonstrate their performance by applying them to biological networks. Further, we illustrate possible use cases using topological network descriptors for performing a quantitative analysis of biological networks. The section *Summary and Outlook* concludes and summarizes the paper and outlines future developments.

Implementation

We implemented a selection of topological network descriptors discussed in [15,27]. Table 1 gives an overview about all implemented network measures with the name of the function to call the corresponding descriptors in R. For a detailed description of all implemented descriptors in QuACN, see the package vignette or additional literature [24,27].

The measures can be categorized within the following groups:

Descriptors based on distances in a graph

This class contains measures that use distances between nodes to capture the structural complexity of the underlying network. A famous and classical representative of this group is the Wiener index [36] that has been defined by the sum of all distances within the network. We also integrated a group of basic distance-based descriptors introduced by Skorobogatov and Dobrynin [37].

Descriptors based on other graph invariants

The descriptors in this class use other graph invariants than distances (e.g. degree, number of vertices, number of edges, etc.) to characterize the structural complexity of complex biological networks. For example, the Zagreb group indices [38] are based on the degree of the vertices. The normalized edge complexity [39] is calculated by using the adjacency matrix and the number of vertices.

Information measures

For an extensive overview of measures of this class, see [16,20,27].

• *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 due to Shannon [19]. The topological information content introduced by Rashevsky [14] and reformulated by Trucco [21] is based on partitions of vertices that are in the same vertex orbit, to calculate the entropy of a graph. Additionally, Mowshowitz [19] investigated mathematical properties of the index to characterize product graphs and other sophisticated measures such as the chromatic information content of a graph.

• *Parametric graph entropy measures*

Measures of this class [27,34] assign a probability value to each vertex of a graph, using so-called information functionals (IFs) which capture structural information of the network. A special information functional quantifies the structural information by using the cardinalities of the corresponding j -spheres [34]. The derived probability distribution is used to calculate the entropy, which has been called Dehmer entropy [34].

As mentioned above, it is not the aim of this manuscript to describe all descriptors in detail. For a better understanding of the used descriptors see the vignette of QuACN and the extensive work of Dehmer and Mowshowitz [27] on information measures for networks.

Table 1 Overview about the implemented topological network descriptors

Name	Symbol	R function	Ref.
Descriptors based on distances			
Skorobogatov indices	$D(G)$	dobrynin(g)	[37]
Wiener index	$W(G)$	wiener(g)	[36]
Hararay index	$H(G)$	harary(g)	[53]
Balaban J index	$J(G)$	balabanJ(g)	[54]
Compactness	$C(G)$	compactness(g)	[55]
Product of row sums index	$PRS(G)$	productOfRowSums(g)	[56]
Hyper-distance-path index	$D_p(G)$	hyperDistancePathIndex(g)	[24]
Descriptors based on other invariants			
Index of total adjacency	$A(G)$	totalAdjacency(g)	[39]
Zagreb group indices 1	$Z_1(G)$	zagreb1(g)	[38]
Zagreb group indices 2	$Z_2(G)$	zagreb2(g)	[38]
Randić index	$R(G)$	randic(g)	[57]
The complexity index B	$B(G)$	complexityIndexB(g)	[39]
Normalized edge complexity	$E_N(G)$	normalizedEdgeComplexity(g)	[39]
Classical entropy-based descriptors			
Topological information content	$I_{orb}^V(G)$	topologicalInfoContent(g)	[14,21]
Bonchev-Trinajstić index 1	$I_D(G)$	bonchev1(g)	[42]
Bonchev-Trinajstić index 2	$I_D^W(G)$	bonchev2(g)	[42]
BERTZ complexity index	$C(G)$	bertz(g)	[58]
Radial centric info index	$I_{c,r}(G)$	radialCentric(g)	[20]
Vertex degree equality-based ii.	$I_{deg}(G)$	vertexDegree(g)	[20]
Balaban-like information index U	$U(G)$	balabanlike1(g)	[40]
Balaban-like information index X	$X(G)$	balabanlike2(g)	[40]
Graph vertex complexity index	$I_V(G)$	graphVertexComplexity(g)	[59]
Dehmer entropy with information functionals using			
the j -spheres	$I_{f^V}(G)$	infoTheoreticGCM(g,infofunc="sphere")	[43]
path lengths	$I_{f^P}(G)$	infoTheoreticGCM(g,infofunc="pathlength")	[43]
vertex centrality	$I_{f^C}(G)$	infoTheoreticGCM(g,infofunc="vertcent")	[43]
degree-degree associations	$I_{f^\Delta}(G)$	infoTheoreticGCM(g,infofunc="degree")	[49]

This table gives an overview about the implemented topological network descriptor including the function name in QuACN and the reference to the corresponding publication.

QuACN is entirely written in R and detailed help is available according to the R documentation standards.

Results

The examples below show the functionality of QuACN by using a selection of small example graphs, which are shown by Figure 2. Our goal is to show how the methods work and to apply the measures to a multitude of complex networks that may lead to novel applications in the field.

Example Graphs

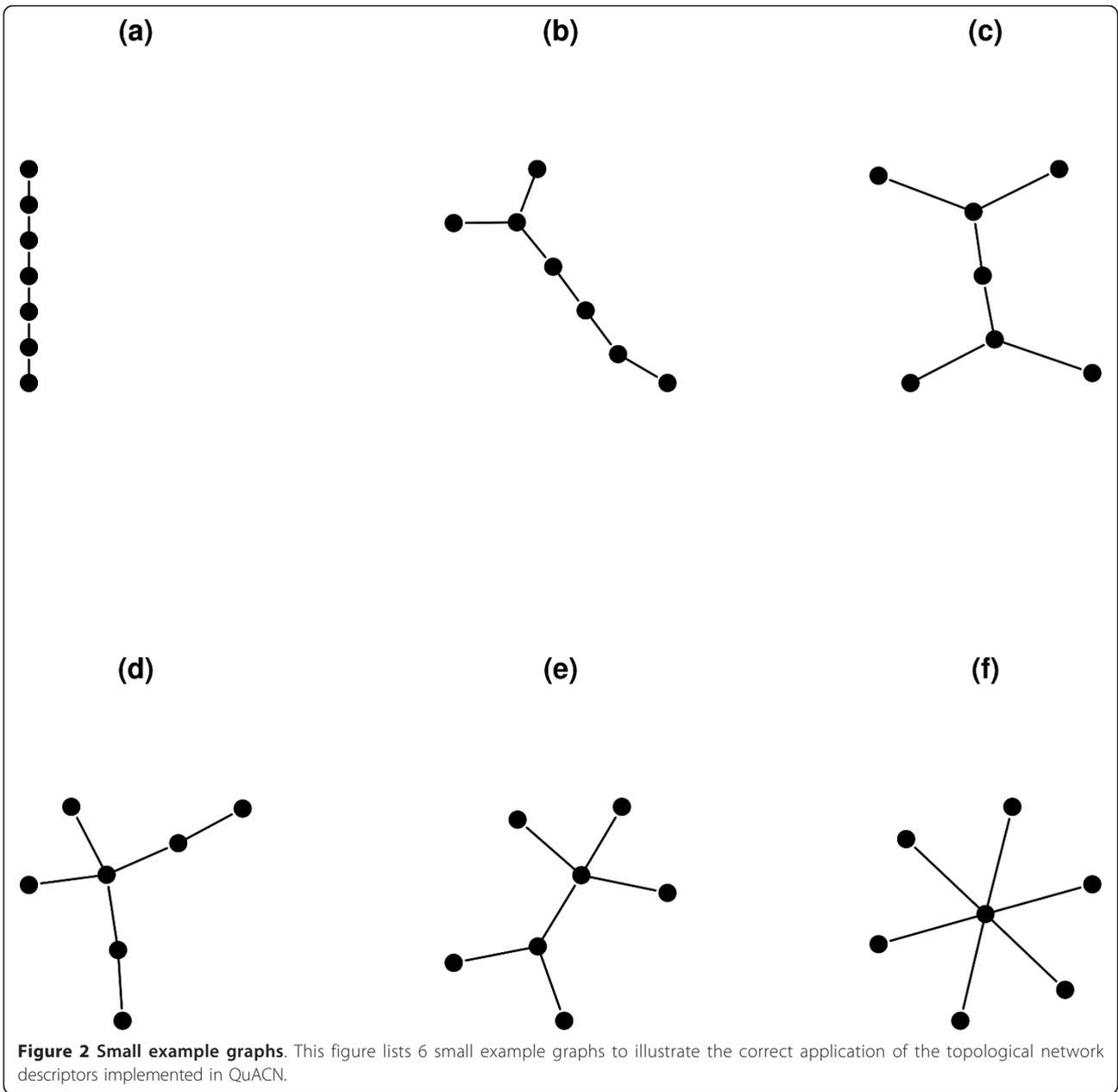
To demonstrate the usefulness of topological network descriptors, we consider Figure 2, showing six undirected example graphs. An undirected graph or network $G = (V, E)$ consists of a non-empty vertex set V . E is

called the edge set of G and is the set of unordered pairs of elements of V . We calculate exemplary a set of descriptors consisting of the Wiener index $W(G)$ [36], the Balaban-like index $X(G)$ [40], the topological information content $I_{orb}(G)$ [14,21] and the Dehmer entropy $I_{f^V}(G)$ [34]. The results are shown in Table 2.

Calling the corresponding methods in R can be done in different ways. The following example shows how to calculate the Wiener index from the `graphNEL`-object g , representing the example graph (a) in Figure 2.

```
> wiener(g)
[1]56
```

As all descriptors are implemented as R-functions it is possible to easily calculate them for a set of graphs using the methods from the `apply`-family.



```
> sapply(glist,balabanlike2)
0.5978703 0.6932045 0.8190124
(a) (b) (c) (d) (e) (f)
```

Table 2 Selected descriptors for the small example graphs

	(a)	(b)	(c)	(d)	(e)	(f)
Wiener index W	56.0000	52.0000	48.0000	44.0000	42.0000	36.0000
Balaban-like index X	0.5979	0.6932	0.8190	1.0492	1.1452	1.8204
Topological information content I_{orb}	1.9502	2.5216	1.3788	1.9502	1.8424	0.5917
Dehmer entropy I_{fv}	2.7648	2.7533	2.7432	2.7282	2.7305	2.7391

Results of some selected descriptors applied to the small example graphs shown in Figure 2.

1.0491707 1.1451745 1.8204321

Note that each descriptor has at least two parameters as listed in Table 3. However, passing the distance matrix to the corresponding function is optional. If the parameter remains empty or is set to *NULL* the distance matrix will be calculated within each function. If calculating more than one descriptor for one graph, it is recommended to calculate the distance matrix separately and pass it to each method, instead of recalculating it again. Particularly when using large networks it can save a lot of time to calculate the distance matrix only once. It will enhance the performance of the calculations significantly. We demonstrate the pre-calculation of the distance matrix in the next example, where we calculate four descriptors for the example graphs in Figure 2. The results of the below listed function call are listed in Table 2.

```
> descriptors <- sapply(glist, function(g){
+   dm <- distanceMatrix(g)
+   result = list()
+   result[["Wiener"]] <- wiener(g, dist = dm)
+   result[["BalabanLike2"]] <- balabanlike2(g, dist =
dm)
+   result[["topologicalInfoContent"]] <-
+     topologicalInfoContent(g, dist = dm)$Iorb
+   result[["Dehmer_jsphere"]] <-
+     infoTheoreticGCM(g,
+       dist = dm,
+       coeff="exp",
+       infofunct="sphere",
+       lambda = 1000)$entropy
+   return(result)
+ })
```

Calling topological information content [14,19,21] and the Dehmer entropy [34] returns a list of different variables. In the example we only use the entropy value of the descriptor. The call of the function works like all other methods, but it returns a list of different values. To explain the result of this function we apply it to graph (c) in Figure 2:

```
> topologicalInfoContent(glist[[3]])
$entropy
[1]1.378783
$orbitals
[1]4 2 1
```

The implementation of the topological information content returns a list containing the entropy (*\$entropy*) and the number of nodes within the same orbit (*\$orbitals*). This information can be used for different other applications, e.g. to determine a graph prototype, see [41].

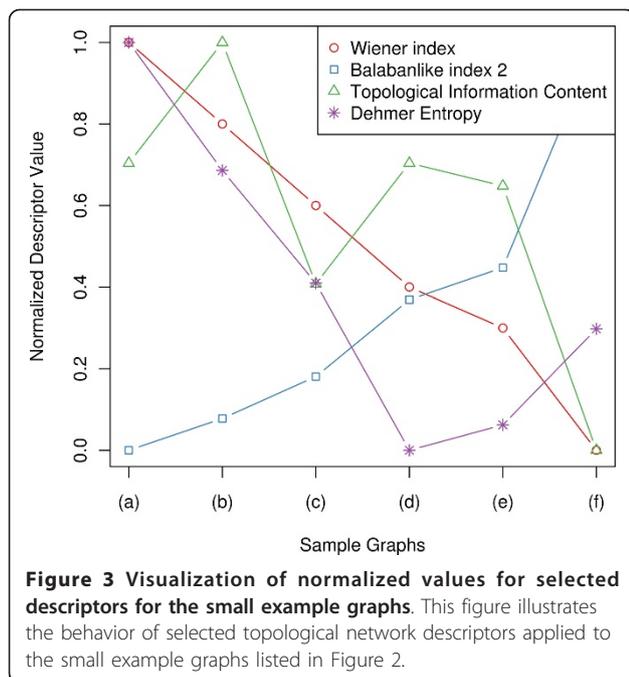
The numerical results of the foregoing example can be seen in Table 2. The visual representation of the normalized results in Figure 3 shows the different behavior of the topological network descriptors using the example graphs. The example graphs start with a linear graph (a) and the branching of the graphs increases towards (f). In this context, branching correlates with the number of terminal vertices (endvertices) [42]. The Wiener index is known as an index to detect molecular branching [24], and one can see that the Wiener index represents increasing branching with decreasing values. Furthermore we can see in this example, that the Balaban-like index $X(G)$ also detects branching well. Note, that its values are just given in a reverse order. The topological information content is based on partitions of vertices that are in the same vertex orbit. But calculating I_{orb} shows that the quantity does not reflect branching properly. As known, I_{orb} is a symmetry-based measure rather than an index for structural complexity [27]. In this example, the Dehmer entropy with monotonously decreasing weighting parameter c_i and the information functional using the j -spheres, neither reflects branching appropriately. The information functional using the j -spheres [34] itself has been used to investigate the information spread in a network [43,44]. However, with a different parameter setting, the Dehmer entropy reflects branching of certain networks meaningfully [45].

However, this simple but demonstrative example indicates that not every topological information index is suitable for a particular problem. It is a challenging task to derive general statements about the structural complexity captured by such measures [15]. It is even harder to connect biological properties with topological network descriptors. Despite the fact that we often do not know the biological interpretation of topological network measures exactly, they can be helpful in a broad range of biological questions. For example, classifying biological data or identifying meaningful biological features, by analyzing the topology of biological networks.

Table 3 Common parameters for each function in QuACN

Name	Type	Description	Mandatory
g	graphNEL	The graph that represents the network.	yes
dist	matrix	The distance matrix of g. If this parameter remains empty or is set to NULL, the distance matrix will be calculated separately within the corresponding R-function.	no

This table shows the two parameters that are common for every method.



To conclude this section, we want to emphasize that one has to understand the selected descriptors and measures in detail to interpret the results correctly. Note that topological network analysis is a non-trivial task and one has to know specific properties of the descriptors to solve a particular problem dealing with networks. One example is the group of Balaban-like indices $X(G)$ and $U(G)$. For a graph with two vertices connected with one edge the index is defined as infinite. That is also returned by the QuACN-method but accompanied by a warning:

```
> g = new("graphNEL")
> # add nodes
> g = addNode("1",g)
> g = addNode("2",g)
> g = addEdge("1","2",g,1)
> balabanlike1(g)
```

[1]Inf

Warning message:

In balabanlike1(g): Graphs with
 $|V| < 3$ result in: Inf!

It is important to know how the different descriptors are defined, when processing and interpreting the results. Note, that not each combination of networks and descriptors could be tested and considered within the exception handling. Keep in mind that applying QuACN to concrete networks can result in special values (i.e.: infinite (Inf), not available (NA) or not a number (NaN)).

The next section shows an example of a possible application of QuACN with biological networks. We

will also use this chapters to explain the usage of more complex descriptors implemented in QuACN.

Supervised Machine Learning for Prostate Cancer Networks

In this section, we present an application of topological network descriptors to classify gene networks inferred from gene expression data. Note, we do not aim to justify network-based approaches itself and compare them to alternative approaches. In fact, a large body of literature dealing with networks does exist, i.e., see [2,4,16].

This example was chosen to explain a possible application of topological network descriptors on biological data. Therefore, we will focus on the methodical usage of the network measures and not on the biological interpretation of the results.

To perform our analysis, we selected seven public available studies of prostate cancer from NCBI GEO and EBI Arrayexpress and inferred networks using the C3NET inference method [5]. This resulted in seven networks $\{G_i^B\}_{i=1}^7$ representing benign tissue (from the control group) and seven networks $\{G_i^C\}_{i=1}^7$ representing cancer tissue. Then we extracted subgraphs from these networks based on the gene ontology (GO) database [46]. For each network and each GO-term we extracted one subgraph containing the genes associated with this specific GO-term. This resulted in a total of 159 networks representing benign tissue and 108 networks representing cancer tissue. The numbers are different because the network structure of G_i^B and G_i^C is different and, hence, not all pathways are captured by these networks. Whenever a subnetwork contained less than 10 genes associated with a GO-term, we excluded this pathway from the analysis. The obtained network sets can be seen as an approximation of two populations. One population represents benign and the second cancerous molecular interactions.

Additionally, we calculated all topological network descriptors available in QuACN, as feature vectors for each of these networks. Afterwards, we performed feature selection and classification using random forest with 10-fold cross-validation (CV). In order to correct the selection bias, an external cross validation is applied to the selection process [47]. In particular, we performed the selection process within each CV-loop [48]. We trained the classifier to classify cancer networks versus benign networks, what lead to a mean classification performance of a F-score of 0.80 and an accuracy of 0.74. This demonstrates that the topological network descriptors, integrated in QuACN, are able to capture group specific structural features meaningfully to distinguish between networks representing prostate cancer and benign tissue. Importantly, this result is not trivial

as one could easily show by using other measures or only a particular fraction thereof, the classification task would result in a random classification. Hence, this result would not be feasible in practice.

As already mentioned we won't focus on a biological representation of the results, as it is the aim of this publication to discuss the methodical perspective of the presented R-package.

One of the measures that showed a significant group effect was the Dehmer Entropy [43]. The Dehmer entropy is a complex measure with several parameters. It is possible to choose the information functional $f(v_i)$, the weighting parameter c_i and the scaling constant λ [49]. The means of these parameters has been discussed in [43]. The user can specify four different information functionals using j -spheres, path lengths, vertex centrality or degree-degree associations [43,49]. We implemented different pre-settings for the weighting parameter c_i : *constant*, *linear*, *quadratic* or *exponential*. A customized setting for c_i can also be declared. The following example shows how to call the function to calculate a Dehmer entropy. The information functional using j -spheres with an exponential setting for c_i and a scaling constant $\lambda = 2500$ are used.

```
> infoTheoreticGCM(gl[[3]], infofunct="sphere",  
+   coeff="exp", lambda = 2500)  
$entropy
```

```
[1]2.743221
```

```
$distance
```

```
[1]160.3339
```

```
$pis
```

```
1 2 3 4
```

```
0.1057720 0.1952924 0.1863273 0.1952924
```

```
5 6 7
```

```
0.1057720 0.1057720 0.1057720
```

```
$fvis
```

```
1 2 3 4
```

```
7.882673 14.554200 13.886071 14.554200
```

```
5 6 7
```

```
7.882673 7.882673 7.882673
```

This function returns a list containing a more comprehensive result than the other measures. Certainly, the list contains the Dehmer entropy denoted by \$entropy. The list entry named \$distance contains the distance of

the entropy from maximum entropy [43]. It also returns the results of calculating the information functional (\$fvis) and the corresponding probability distribution (\$pis). The probability distribution can later be used for further analysis, i.e. estimating the graph prototype of a set of networks [41].

Conclusion

The freely available open source R-package QuACN contains a selection of topological network descriptors. The aim of this manuscript was to explain, how to apply the implemented descriptors correctly to complex biological networks using R. To provide a basic understanding of the application we demonstrated the behavior of the indices by applying them to small example networks. Moreover, we presented an application for supervised machine learning from biological networks by using topological network descriptors. Within these examples we demonstrated the correct usage of the methods included in QuACN. Machine learning is not the only application that topological network descriptors can be used for. They also can be utilized to compare networks. In this sense, Kugler et. al. [41] calculated the Kullback-Leibler divergence to perform an integrative network analysis.

Topological network descriptors have been standard methods in the field of quantitative structure property activity relationship (QSAR/QSPR) [22,34]. The methods implemented in QuACN had already been used for QSAR/QSPR applications, see [22,34]. Further applications of information-theoretic measures had been discussed by Dehmer and Mowshowitz [27].

The indices integrated in QuACN can also be efficiently applied on large networks as their calculation requires polynomial time complexity. However, there also exist some indices whose algorithms are NP-complete (e.g., descriptors based on the subgraph isomorphism problem [50] or the Hosoya index [51]), but they have not been integrated in the package. Importantly, not every index is suitable for any application in network biology and it strongly depends on the underlying research question which measures can be considered as appropriate.

Using the concept of advanced network descriptors is relatively new in systems biology. Advanced network descriptors are able to quantify specific topological characteristics of the underlying network but the interpretation of the structural properties of the applied measures is still an ongoing task [15]. However, modeling biological systems as networks had become an important task in recent systems biology research and created a need for methods to analyze them structurally. Therefore, the topological network measures provided by QuACN can stimulate the research in this field. However, a thorough

analysis to investigate the behavior of topological information indices on biological networks is planned to be performed.

As future work, we plan to apply the integrated measures on various biological research questions, and to extend the range of functions with new promising descriptors for coming versions of QuACN. The next step is to integrate a group of already existing polynomial-based descriptors [22,52]. Finally, we are convinced that this package will turn out to be useful for a community dealing with network biology [16].

Availability and requirements

Project name: QuACN - Quantitative Analysis of Complex Networks

Project home page: <http://cran.r-project.org/web/packages/QuACN/>

Operating system(s): Platform independent

Programming language: R (<http://www.r-project.org>)

License: LGPL

Acknowledgements

Matthias Dehmer thanks the Austrian Science Funds for supporting this work (project P22029-N13). This work was also partly supported by the Tiroler Wissenschaftsfonds and the Standortagentur Tirol (Tiroler Zukunftsstiftung). We thank Matthias Wieser and Andreas Dander who helped to develop the R package.

Author details

¹Institute for Bioinformatics and Translational Research, Department of Biomedical Sciences and Engineering, University for Health Sciences, Medical Informatics and Technology (UMIT), EWZ 1, Hall in Tirol, Austria.

²Computational Biology and Machine Learning Lab, Center for Cancer Research and Cell Biology, School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, 97 Lisburn Road, Belfast, BT9 7BL, UK.

Authors' contributions

LAJM and KGK implemented and tested the R-package, performed the analysis and interpreted the results. LAJM, KGK, AG, FES and MD wrote the manuscript. MD supervised the study. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 5 August 2011 Accepted: 24 December 2011

Published: 24 December 2011

References

1. Kitano H: Systems Biology: A Brief Overview. *Science (New York, NY)* 2002, **295**:1662-4.
2. Emmert-Streib F, Glazko GV: Network Biology: A Direct Approach to Study Biological Function. *Wiley Interdisciplinary Reviews. Systems biology and medicine* 2010, 1-27.
3. Palsson B: *Systems Biology: Properties of Reconstructed Networks* Cambridge University Press; 2006.
4. Barabási AL, Oltvai ZN: Network Biology: Understanding the Cell's Functional Organization. *Nature Reviews Genetics* 2004, **5**(2):101-13.
5. Altay G, Emmert-Streib F: Inferring the Conservative Causal Core of Gene Regulatory Networks. *BMC Systems Biology* 2010, **4**:132.
6. Meyer PE, Lafitte F, Bontempi G: minet: A R/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information. *BMC Bioinformatics* 2008, **9**:461.
7. Adourian A, Jennings E, Balasubramanian R, Hines WM, Damian D, Plasterer TN, Clish CB, Stroobant P, McBurney R, Verheij ER, Bobeldijk I, van der Greef J, Lindberg J, Kenne K, Andersson U, Hellmold H, Nilsson K, Salter H, Schuppe-Koistinen I: Correlation Network Analysis for Data Integration and Biomarker Selection. *Molecular BioSystems* 2008, **4**(3):249-259.
8. Meyer PE, Marbach D, Roy S, Kellis M: Information-Theoretic Inference of Gene Networks Using Backward Elimination. *Conference on Bioinformatics & Computational Biology (BIOCOMP'10), Las Vegas/USA* 2010, II:700-705.
9. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A: ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics* 2006, **7** Suppl 1((Suppl 1)):S7.
10. Langfelder P, Horvath S: WGCNA: An R Package for Weighted Correlation Network Analysis. *BMC Bioinformatics* 2008, **9**:559.
11. Oppen-Rhein R, Strimmer K: From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Systems Biology* 2007, **1**:37.
12. Højsgaard S: The mimR Package for Graphical Modelling in R. *Journal of Statistical Software* 2004, **11**:6:1-13.
13. Scutari M: Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software* 2010, **35**(3):1-22.
14. Rashevsky N: Life, Information Theory, and Topology. *Bulletin of Mathematical Biophysics* 1955, **17**:229-235.
15. Dehmer M, Barbarini N, Varmuza K, Graber A: A Large Scale Analysis of Information-Theoretic Network Complexity Measures Using Chemical Structures. *PLoS ONE* 2009, **4**(12).
16. Emmert-Streib F, Dehmer M: Networks for Systems Biology: Conceptual Connection of Data and Function. *IET Systems Biology* 2011, **5**(3):185-207.
17. Mueller LA, Kugler KG, Dander A, Graber A, Dehmer M: QuACN: An R Package for Analyzing Complex Biological Networks Quantitatively. *Bioinformatics* 2011, **27**:140-141.
18. Dehmer M, Borgert S, Emmert-Streib F: Entropy Bounds for Hierarchical Molecular Networks. *PLoS ONE* 2008, **3**(8):e3079.
19. Mowshowitz A: Entropy and the Complexity of the Graphs I: An Index of the Relative Complexity of a Graph. *Bulletin of Mathematical Biophysics* 1968, **30**:175-204.
20. Bonchev D: *Information Theoretic Indices for Characterization of Chemical Structures* Research Studies Press, Chichester; 1983.
21. Trucco E: A Note on the Information Content of Graphs. *Bulletin of Mathematical Biology* 1956, **18**(2):129-135.
22. Dehmer M, Sivakumar L, Varmuza K: Uniquely Discriminating Molecular Structures Using Novel Eigenvalue Based Descriptors. *MATCH Communications in Mathematical and in Computer Chemistry* 2012, **67**:147-172.
23. Bonchev D, Mekenyan O, Trinajstić N: Isomer Discrimination by Topological Information Approach. *Journal of Computational Chemistry* 1981, **2**(2):127-148.
24. Todeschini R, Consonni V, Mannhold R: *Handbook of Molecular Descriptors* Wiley-VCH [Weinheim, Germany]; 2002.
25. Xia K, Fu Z, Hou L, Han J: Impacts of Protein-Protein Interaction Domains on Organism and Network Complexity. *Genome Research* 2008, **18**(9):1500.
26. Mazurie A, Bonchev D, Schwikowski B, Buck G: Phylogenetic Distances are Encoded in Networks of Interacting Pathways. *Bioinformatics* 2008, **24**(22):2579.
27. Dehmer M, Mowshowitz A: A History of Graph Entropy Measures. *Information Sciences* 2011, **181**:57-78.
28. Liu K, Feng J, Young S: PowerMV: A Software Environment for Molecular Viewing, Descriptor Generation, Data Analysis and Hit Evaluation. *Journal of Chemical Information and Modeling* 2005, **45**(2):515-522.
29. Wegner J, Zell A: JOELib: A Java Based Computational Chemistry Package. *6th Darmstädter Molecular-Modelling Workshop* 2002.
30. Todeschini R, Consonni V, Mauri A, Pavan M: Software Dragon: Calculation of Molecular Descriptors, Department of Environmental Sciences. *Talente* 2003.
31. Lee SK, Lee IH, Kim HJ, Chang GS, Chung JE, No KT: The PreADME Approach: Web-Based Program for Rapid Prediction of Physico-Chemical, Drug Absorption and Drug-Like Properties. *euro QSAR 2002 - Designing Drugs and Crop Protectants: Processes Problems and Solutions* 2002.

32. Csardi G, Nepusz T: **The igraph Software Package for Complex Network Research, Complex Systems:1695.** *InterJournal* 2006.
33. Carey V, Long L, Gentleman R: **RBGL: An Interface to the BOOST Graph Library**, [R package version 1.2].
34. Dehmer M, Varmuza K, Borgert S, Emmert-Streib F: **On Entropy-based Molecular Descriptors: Statistical Analysis of Real and Synthetic Chemical Structures.** *Journal of Chemical Information and Modeling* 2009, **49**:1655-1663.
35. In *Analysis of Complex Networks: From Biology to Linguistics*. Edited by: Dehmer M, Emmert-Streib F. Wiley VCH Publishing; 2009.
36. Wiener H: **Structural Determination of Paraffin Boiling Points.** *Journal of the American Chemical Society* 1947, **69**:17-20.
37. Skorobogatov VA, Dobrynin AA: **Metrical Analysis of Graphs.** *MATCH Communications in Mathematical and in Computer Chemistry* 1988, **23**:105-155.
38. Diudea MV, Gutman I, Jäntschi L: *Molecular Topology* Nova Publishing [New York, NY, USA]; 2001.
39. Bonchev D, Rouvray DH: *Complexity in Chemistry, Biology, and Ecology* Mathematical and Computational Chemistry, Springer [New York, NY, USA]; 2005.
40. Balaban AT, Balaban TS: **New Vertex Invariants and Topological Indices of Chemical Graphs Based on Information on Distances.** *Journal of Mathematical Chemistry* 1991, **8**:383-397.
41. Kugler K, Mueller L, Gräber A, Dehmer M: **Integrative Network Biology: Graph Prototyping for Co-Expression Cancer Networks.** *PLoS ONE* 2011, **6**(7):e22843.
42. Bonchev D, Trinajstić N: **Information Theory, Distance Matrix and Molecular Branching.** *Journal of Chemical Physics* 1977, **67**:4517-4533.
43. Dehmer M: **Information Processing in Complex Networks: Graph Entropy and Information Functionals.** *Applied Mathematics and Computation* 2008, **201**:82-94.
44. Dehmer M: **Information-Theoretic Concepts for the Analysis of Complex Networks.** *Applied Artificial Intelligence* 2008, **22**(7):684-706.
45. Dehmer M, Emmert-Streib F: **The Structural Information Content of Chemical Networks.** *Zeitschrift für Naturforschung A* 2008, **63**:155-158.
46. Harris M, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, et al: **The Gene Ontology (GO) database and informatics resource.** *Nucleic acids research* 2004, **32** Database: D258.
47. Ambrose C, McLachlan G: **Selection Bias in Gene Extraction on the Basis of Microarray Gene-Expression Data.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(10):6562.
48. Varma S, Simon R: **Bias in Error Estimation When Using Cross-Validation for Model Selection.** *BMC Bioinformatics* 2006, **7**:91.
49. Dehmer M, Emmert-Streib F, Tsoy Y, Varmuza K: **Quantifying Structural Complexity of Graphs: Information Measures in Mathematical Chemistry.** In *Quantum Frontiers of Atoms and Molecules*. Edited by: Putz M. Nova Publishing; 2011:479-498.
50. Eppstein D: **Subgraph Isomorphism in Planar Graphs and Related Problems.** *Journal of Graph Algorithms and Applications* 1999, **3**(3):1-27.
51. Hosoya H: **Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons.** *Bulletin of the Chemical Society of Japan* 1971, **44**(9):2332-2339.
52. Ellis-Monaghan J, Merino C: **Graph polynomials and their applications I: The Tutte polynomial.** *Structural Analysis of Complex Networks* 2011, **219**-255.
53. Balaban AT, Ivanciuc O: **Historical Development of Topological Indices.** In *Topological Indices and Related Descriptors in QSAR and QSPAR*. Edited by: Devillers J, Balaban AT. Gordon and Breach Science Publishers [Amsterdam, The Netherlands]; 1999:21-57.
54. Balaban AT: **Highly Discriminating Distance-based Topological Index.** *Chemical Physics Letters* 1982, **89**:399-404.
55. Doyle JK, Garver JE: **Mean Distance in a Graph.** *Discrete Mathematics* 1977, **17**:147-154.
56. Schultz HP, Schultz EB, Schultz TP: **Topological organic chemistry. 4. Graph theory, matrix permanents, and topological indices of alkanes.** *Journal of Chemical Information and Computer Sciences* 1992, **32**:69-72.
57. Li X, Gutman I: *Mathematical Aspects of Randić-Type Molecular Structure Descriptors* Mathematical Chemistry Monographs, University of Kragujevac and Faculty of Science Kragujevac; 2006.
58. Bertz SH: **The First General Index of Molecular Complexity.** *Journal of the American Chemical Society* 1981, **103**:3241-3243.
59. Raychaudhury C, Ray SK, Ghosh JJ, Roy AB, Basak SC: **Discrimination of Isomeric Structures Using Information Theoretic Topological Indices.** *Journal of Computational Chemistry* 1984, **5**:581-588.

doi:10.1186/1471-2105-12-492

Cite this article as: Mueller et al.: Structural Measures for Network Biology Using QuACN. *BMC Bioinformatics* 2011 **12**:492.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

