BMC Systems Biology

Poster presentation

Abstract analysis of pathways using the BN++ software framework

Jan Küntzer*[1], Benny Kneissl[1], Oliver Kohlbacher[2] and Hans-Peter Lenhof[1]

Address: [1]Center for Bioinformatics, Saarland University, 66041 Saarbrücken, Germany and [2]Center for Bioinformatics/Wilhelm Schickard Institute for Computer Science, Eberhard-Karls-Universität Tübingen, 72076 Tübingen, Germany

Email: Jan Küntzer* - kuentzer@bioinf.uni-sb.de

* Corresponding author

This abstract is available from: http://www.biomedcentral.com/1752-0509/1?issue=S1

Introduction

Technological advances in high-throughput techniques and efficient data acquisition methods have resulted in a vast amount of life science data, including pathway information such as metabolic and regulatory pathways. The rapid increase of these data for various organisms offers the possibility to perform analyses on the networks for single organisms (intra-species) as well as across different organisms (inter-species). However, the sheer amount and heterogeneity of the data pose a major challenge and call for an integrative system, allowing to manage all this information. With BN++, especially its C++ framework, we presented such a system [2]. In contrast to databases (e.g. KEGG, Reactome, IntAct,...), which offer only predefined analyses such as minimal connected component or pathway detection from a start to an end compound, the mathematical graph representation in BN++ allows additionally the implementation of own routines. The analysis of biochemical pathway information has different applications, e.g. in the process of target identification, drug design and in the search for causes of genetic diseases. Therefore, nodes or edges are removed and alternative pathways in an organism need to be identified. In basic research these networks can be used for the comparison of metabolic processes of different organisms. For example the information on the metabolism of one organism can be used to understand the newly sequenced genome (and, hence the metabolic pathways) of another organism as presented in [1].

Background

To understand the mathematical graph representation, we first need to define some concepts: We define $G(V, E)$ to be a **mathematical graph**, where V denotes a finite set of nodes of G and $E = (V \times V)$ denotes a set of pairs of nodes, called the edges of the G. A graph $G(V, E)$ is defined as **bipartite**, iff $V = V_1 \cup V_2$ can be partitioned into two sets V_1 and V_2 such that $V_1 \cap V_2 = \varnothing$ and $(u, v) \in E = ((V_1 \times V_2) \cup (V_2 \times V_1))$ implies either $u \in V_1$ and $v \in V_2$, or $u \in V_2$ and $v \in V_1$. The modeling of a biochemical pathway as a mathematical graph can be done in different ways, differing in the interpretation of the nodes and edges. Hence, various data models have been developed over the last years. The mostly used models are presented in [3]. We will define three different models in the following: First, we define a **bipartite reaction graph** to be a bipartite graph, where V_1 contains all events and V_2 all compounds. A node A is connected with a directed edge to node B, iff compound A plays the role of an educt in event B or if compound B plays the role of a product in event A. Second, a **compound graph** is defined as a mathematical graph, where the nodes are the chemical compounds. A is connected with a directed edge to B, iff compound A plays the role of an educt and compound B plays the role of a product in the same event. Third, a **event graph** defines a mathematical graph, where the nodes are the events. A node A is connected with a directed edge to B, iff a compound Y plays the role of an educt in the event A and the role of a product in the event B.

Results

The biochemical network library BN++ is a powerful software package for integrating, analyzing, and visualizing biochemical data in the context of networks. The heart of BN++ is built by the comprehensive object-oriented data model BioCore, which allows to model most of the known biochemical processes in metabolic and regulatory pathways. The main concept of BioCore is based on three central classes Event, Role, and Participant. Biochemical processes are modeled as Events with different Participants playing a certain Role. BioCore contains already a huge number of predefined Event classes (Reaction, Interaction, Expression,...), Participant classes (Protein, Gene, DNA, RNA, Compound,...), and Role classes (Maineduct, Sideeduct, Enzyme, Activator, Inhibitor,...). However, it can be easily extended by subclassing from the core classes to model new biochemical knowledge. Numerous databases with different data models and structures have been established. BN++ contains import capabilities for a large number of different external data sources (KEGG, BioCyc, TransPath, DIP, MINT, IntAct, HPRD,...), which can be stored in the data warehouse of the biochemical network library. The C++ framework allows to create a mathematical graph representation for the analysis of complex biochemical data in the context of networks. Since there is no unique mapping of biochemical networks onto a single graph structure, we provide different generic mappings, that enable us to map arbitrary BioCore classes onto the nodes and edges of a graph. We integrated an event graph, a compound graph, as well as a bipartite reaction graph representation of the data into the framework. The edges for enzymatic reactions in the compound graph representation are optionally labelled with the enzyme catalyzing the event. In addition the compound graph is only built from the Mainproducts and Maineducts in the reaction. All the graphs can be generated from the BN++ data warehouse or the BioCore data model by one single line of code. The internal implementation of the graph data structure is based on the Boost Graph Library (BGL). BGL also provides a number of graph algorithms like shortest paths, minimum spanning trees, connected components, etc. Furthermore, we integrated new algorithms like the k-shortest path algorithm. The rapid prototyping library BN++ allows to focus on the analysis without the need neither to deal with an implementation of the import of external data nor with an implementation of a graph representation. To show the power and the ease-of-use of the BN++ software framework, we integrated a pathway similarity algorithm on the basis of the algorithm introduced by Pinter et al. [4]. The algorithm finds for a given pathway P all similar pathways in a biochemical network T. The similarity between the participants of the pattern and the text are defined by a scoring matrix. As a result, the pathways are ranked by a similarity score.

Conclusion

With the Biochemical Network Library BN++ we present a rapid prototyping library, which offers the possibility to analyze complex data in the context of biological networks with little effort. BN++ provides an interface for a large number of external data sources allowing to combine the data from various data models with one single library. The rich functionality of different graph representations gives the possibility to chose a suitable representation for a given application. The library offers a huge variety of different standard routines like single-source-shortest-path, minimum-spanning-tree, strongly-connected-components, etc. These routines can be used to implement own applications as shown in the section above. All the functionality can be easily used by mostly one single line of code using the rapid prototyping capability of BN++. With BN++ a user can focus on his application. The internal graph implementation using the BGL is generic, in the same sense as the Standard Template Library (STL). This results in a high performance and robustness of the system.

Acknowledgements

References

1. Sirava M, Schafer T, Eiglsperger M, Kaufmann M, Kohlbacher O, Bornberg-Bauer E, Lenhof HP: **BioMiner--modeling, analysing, and visualizing biochemical pathways and networks.** *Bioinformatics* 2002, **18(Suppl 2):**S219-S230.
2. Küntzer J, Gerasch A, Backes C, Hildebrandt A, Kaufmann M, Kohlbacher O, Lenhof HP: **BN++ – A Biological Information System.** *Journal of Integrative Bioinformatics* 2006, **3(2):**34.
3. Deville Y, Gilbert D, van Helden J, Wodak SJ: **An overview of data models for the analysis of biochemical pathways.** *Briefings in Bioinformatics* 2003, **4(3):**246-259.
4. Pinter R, Rokhlenko O, Yeger-Lotem E, Ziv-Ukelson M: **Alignment of metabolic pathways.** *Bioinformatics* 2005, **21(16):**3401-3408.