CentiLib: comprehensive analysis and exploration of network centralities

Johannes Gräßler[1], Dirk Koschützki[2] and Falk Schreiber[1,3,*]

[1]Department of Natural Sciences III, Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle 06120, [2]Department of Computer and Electrical Engineering, Furtwangen University of Applied Sciences, Furtwangen 78120, and [3]Leibniz-Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Gatersleben 06466, Germany

Associate Editor: Martin Bishop

ABSTRACT

Summary: CentiLib is a library and plug-in for the comprehensive analysis and exploration of network centralities. It provides 17 different node centrality and four graph centrality measures in a user-friendly interface and supports the exploration of analysis results within the networks. Its architecture allows for easy adaption to Java-based network analysis, simulation and visualization tools, which is demonstrated by providing the plug-in for two popular network analysis tools—Cytoscape and Vanted. With the ability to quantitatively analyze biological networks in an interactive and visual manner, CentiLib supports a better understanding of complex biological networks and processes.

Availability and implementation: Software with manual and tutorials is freely available at http://centilib.ipk-gatersleben.de/

Contact: schreibe@ipk-gatersleben.de

1 INTRODUCTION

The development of various highly parallel techniques for biological research and data generation has led to the advance of methods to analyze this data. Many tools have been developed to visualize, rearrange, combine, edit and analyze biological data. Biological networks, which can be derived from experimental data or downloaded from publicly available databases, are an increasingly important data structure in life science research. Since many investigated networks are very large and elements are connected to each other in many ways, it is still difficult to analyze the structure of such a network and the correlations between elements or groups of elements. Therefore, several tools allow statistical, visual and other analysis methods.

An important measurement to analyze networks, which is widely used in other sciences such as sociology and becoming increasingly significant in biology, is the centrality of nodes. Centrality measures are used to rank elements in a network according to different importance concepts. This can be very helpful, especially for biological networks, because often there are only a few elements crucial to the underlying biological process. Centralities can help to find those elements without time- and cost-intensive experiments, or can at least guide experiments.

There are already some tools available which support the analysis of network centralities. These are either stand-alone tools such as Visone (Brandes and Wagner, 2005), Pajek (Batagelj and Mrvar, 2003), Network Analyzer (Assenov *et al.*, 2008) and CentiBiN (Junker *et al.*, 2006a), or extensions (plug-ins) to other tools such as CentiScaPe (Scardoni *et al.*, 2009) for Cytoscape. CentiBiN is the stand-alone tool which offers most centralities. It provides 17 node centralities but the user has to install and become accustomed to new software, and further analysis methods and functions such as importing and analyzing experimental data are absent. A general disadvantage of most stand-alone tools is that the biological context is absent; hence one has to export the results and import them in a biology-focused tool to compare and combine them with other experimental data. CentiScaPe is a plug-in for Cytoscape and therefore provides new functionality to a well-known tool for biological network analysis. However, it is only available for Cytoscape and it provides only nine centrality measures.

There are more stand-alone tools and extensions available but all of them have similar disadvantages as those mentioned above. To overcome these drawbacks and to make centrality measures accessible to software tools which are already employed by users, CentiLib bundles 17 node and four graph centralities in a Java library, containing a ready-to-use user interface which can be integrated into nearly every Java-based tool that works with graphs and uses a plug-in mechanism. We provide this integration for two tools, Cytoscape (Shannon *et al.*, 2003) and Vanted (Junker *et al.*, 2006b), , see Figure 1 for a screenshot.

Besides the integration into existing tools, software developers are able to use CentiLib as a pure Java library to compute all centralities via a single and simple to use class. Furthermore, CentiLib not only allows the computation of centralities for directed and undirected graphs but also for weighted ones.

2 METHODS AND IMPLEMENTATION

CentiLib is an easy-to-use library and user-friendly plug-in for the analysis and visual exploration of centralities in networks. It is implemented in Java (requires Java 1.6 or later) and employs the open source libraries JUNG2 (http://jung.sourceforge.net/) for graphs and Colt (http://acs.lbl.gov/software/colt/) for matrix operations. CentiLib focuses on efficient computation of centralities, for example, the centralities betweenness, closeness, eccentricity, radiality and stress use the fast algorithm of Brandes (Brandes, 2001) and are implemented in such a way that all of them can be computed at once.

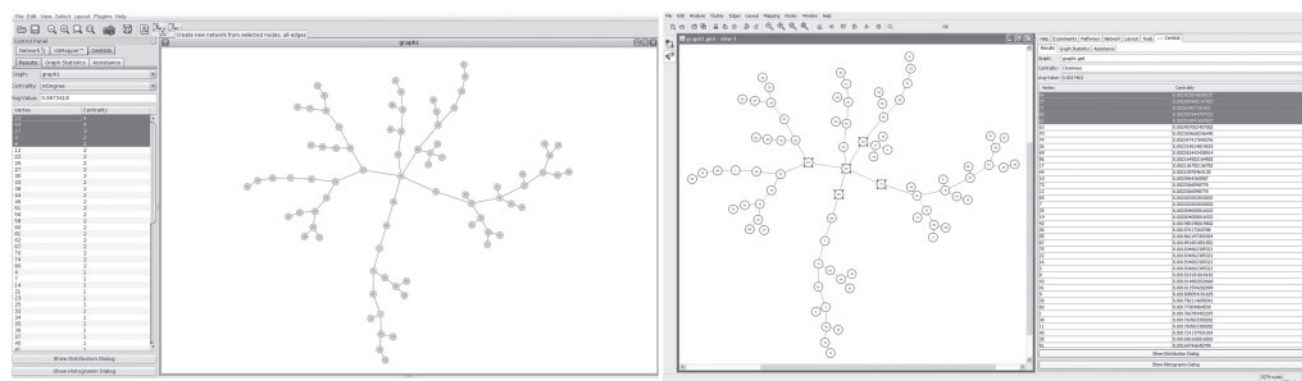*To whom correspondence should be addressed.

Fig. 1. Screenshots of CentiLib for Cytoscape (left) and Vanted (right).

2.1 Centrality measures and exploration

CentiLib not only enables the computation of various centralities but also the exploration of those values. The user can select elements in the graph to see their ranks and values in a result list or select elements from the result list to mark the appropriate graph elements. Once the most interesting elements are selected, different centralities can be compared by switching through the list of already computed ones.

Further analysis with other tools is possible by exporting the centrality values in .vec or .tsv format or saving them as node attributes in the graph. The following node centralities are available: degree, in-/out-degree, eccentricity, closeness, centroid, stress, radiality, current-flow closeness, Katz status index, shortest-path betweenness, current-flow betweenness, eigenvector, hits-hubs/-authorities, Hubbel index and Page-rank.

2.2 Library and plug-in

CentiLib aims to be reusable in an easy way for many different tools. Therefore, CentiLib makes intensive use of Java interfaces. That is, the generic JUNG2-interface Graph<V,E> must be implemented, and the graph structure in question then wrapped to use the library. This way CentiLib is independent of the data structures used and no further changes are required. After wrapping the graph, centralities can be computed via the class *CentralityHandler*.

The integration of the contained ready-to-use user interface is done similarly. Usually, plug-ins are commonly implemented for a single application and therefore access the applications data structures directly and communicate or interact with the application. Instead of this traditional approach, CentiLib uses four interfaces to access the graphs, node names and attributes; and one interface to interact with the application. Thus, an adoption to another target application requires only the need to define those interfaces.

2.3 User interface provided to Cytoscape and Vanted

The CentiLib user interface has so far been integrated as a plug-in into Cytoscape and Vanted. It contains the panels *Assistance*, *Graph statistics* and *Results*. To compute centralities, the user only needs to select, in the *Assistance* panel, one or more graphs from a list, decide which centralities they want to compute, and start the process. The results can be investigated in the *Results* panel. By selecting the required graph and centrality, the user gets a list with all nodes in descending order of their centrality value. List elements can be marked in the graph by selecting them in the result list and vice versa. Because centrality values are only relative, the average value for a given centrality is available too. To use computed values in other programs or keep them for later sessions, they can be exported to files or added as node attributes. Plotting the overall results as a histogram or distribution dialog is also possible. To help users understanding a network, the *Graph statistics* panel gives some information about a graph such

as simplicity, connectedness, loop-freeness, average distance, clustering coefficient, diameter and wiener index. The *Graph statistics* panel also allows edge weights selection. A dropdown list contains all available numeric edge attributes. By selecting one of them, the appropriate values will be utilized as edge weights in the computation of the centrality values. Another feature is the location and extraction of connected or strongly connected components, as some centralities can only be computed for (strongly) connected graphs. Therefore, it is possible to create new graphs by extracting these components.

3 DISCUSSION

CentiLib is a library and graphical user interface that can be easily adapted to many Java-based network analysis tools which provide a plug-in mechanism. It provides 17 node centrality measures and four graph centrality measures, covering a comprehensive set of centralities. CentiLib is fast, user-friendly and allows not only the computation and export of centralities, but also their exploration in the network context. In addition to the library, two adapted versions for Cytoscape and Vanted are provided. In contrast to many other tools for the computation of centralities, CentiLib allows the computation of centrality values also for weighted graphs. We believe that the easy access to many established centrality measures embedded in commonly used software tools will provide scientists from the life sciences with new possibilities for meaningful analysis of biological networks.

Conflict of Interest: none declared.

REFERENCES

Assenov,Y. *et al.* (2008) Computing topological parameters of biological networks. *Bioinformatics*, **24**, 282–284.

Batagelj,V. and Mrvar,A. (2003) Pajek—analysis and visualization of large networks. In Jünger,M. and Mutzel,P. (eds) *Graph Drawing Software*, Springer, Berlin, Heidelberg, pp. 77–103.

Brandes,U. (2001) A faster algorithm for betweenness centrality. *J. Math. Sociol.*, **25**, 163–177.

Brandes,U. and Wagner,D. (2005) Visone—analysis and visualization of social networks. In Jünger,M. and Mutzel,P. (eds) *Graph Drawing Software*, Springer, Berlin, Heidelberg, pp. 321–340.

Junker,B.H. *et al.* (2006a) Exploration of biological network centralities with CentiBiN. *BMC Bioinformatics*, **7**, 219.

Junker,B.H. *et al.* (2006b) Vanted: a system for advanced data analysis and visualisation in the context of biological networks. *BMC Bioinformatics*, **7**, 109.

Scardoni,G. *et al.* (2009) Analyzing biological network parameters with CentiScaPe. *Bioinformatics*, **25**, 2857–2859.

Shannon,P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.