CentiLib: comprehensive analysis and exploration of network centralities

Johannes Gräßler1, Dirk Koschützki2 and Falk Schreiber1,3,*

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

Received on October 26, 2011; revised on February 10, 2012; accepted on February 27, 2012

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 Vison (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 Cytoscape. 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 is an easy-to-use library and user-friendly plug-in for the analysis and exploration of centralities in networks. It is downloadable 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 Vison (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 Cytoscape. 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.
CentiLib makes intensive use of Java interfaces. That is, the distribution dialog is also possible. To help users understanding a network, or added as node attributes. Plotting the overall results as a histogram or programs or keep them for later sessions, they can be exported to files vice versa. Because centrality values are only relative, the average value gets a list with all nodes in descending order of their centrality value. List 2.3 User interface provided to Cytoscape and Vanted

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, centrod, 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 GraphVIE 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 approach, CentiLib uses four interfaces to access the graphs, node names and attributes; and one interface to interact with the application. Instead of this traditional approach, 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, centrod, 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 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 embodied in commonly used software tools will provide scientists from the life sciences with new possibilities for meaningful analysis of biological networks.

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 embodied 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


