

CET: A Tool for Creative Exploration of Graphs

Stefan Haun¹, Andreas Nürnberger¹, Tobias Kötter²,
Kilian Thiel², and Michael R. Berthold²

¹ Data and Knowledge Engineering Group,
Faculty of Computer Science, Otto-von-Guericke-University, Germany
<http://www.findke.ovgu.de>

² Nycomed-Chair for Bioinformatics and Information Mining
University of Konstanz, Germany
<http://www.inf.uni-konstanz.de/biomi1>

Abstract. We present a tool for interactive exploration of graphs that integrates advanced graph mining methods in an interactive visualization framework. The tool enables efficient exploration and analysis of complex graph structures. For flexible integration of state-of-the-art graph mining methods, the viewer makes use of the open source data mining platform KNIME. In contrast to existing graph visualization interfaces, all parts of the interface can be dynamically changed to specific visualization requirements, including the use of node type dependent icons, methods for a marking if nodes or edges and highlighting and a fluent graph that allows for iterative growing, shrinking and abstraction of (sub)graphs.

1 Introduction

Today's search is still concerned mostly with keyword-based searches and the closed discovery of facts. Many tasks, however, can be solved by mapping the underlying data to a graph structure and searching for structural features in a network, e.g. the connection between certain pages in Wikipedia¹ or the environment of a specific document. Exploring a hyperlink structure in a graph representation enables these tasks to be fulfilled much more efficiently. On the other hand, graph visualization can handle quite large graphs, but is rather static, i.e. the layout and presentation methods calculate the graph visualization once and are well suited for interactions, such as adding or removing nodes. One of the famous graph layout methods, the *Spring Force Layout*, can yield very chaotic results when it comes to small changes in the graph, leading to a completely different layout if just one node is removed. Since a user relies on the node positions during interaction with the graph, such behavior is not desirable.

With the *Creative Exploration Toolkit (CET)*, we present a user interface with several distinct features:

- support of interactive graph visualization and exploration,
- integration of a modular open source data analytics system,

¹ <http://www.wikipedia.org>

- easy configuration to serve specific user requirements.

In the following sections, we will describe these features in more detail.

2 The Creative Exploration Toolkit

The Creative Exploration Toolkit (CET) is the user interface that visualizes the graph and allows interaction. The global design, shown in Figure 1, consists of

- a *dashboard* at the top, where the controls are located,
- a *logging area*, below, to show information on running processes and the tool status,
- a *sidebar* on the right which displays detailed information about a node,
- and the *workspace* in the center, which is used for visualization.

We currently use the *Stress Minimization Layout*[3] to determine the initial graph layout, which enables the user to interact with the graph: Nodes can be moved to create certain arrangements, nodes can be selected, and nodes can be expanded by double-clicking them. Additionally, the user may issue keyword-based queries. The corresponding results consists of graphs and can be visualized as well. Subsequent query results are added to the graph, enabling the user to explore the graph itself and the structures between the query results.

While CET takes care of graph visualization and presentation, special semantics are not supported. For example, the shortest path between two nodes is displayed by highlighting all nodes on the path. However, the user interface is not aware of the path-property, but only displays the highlight attribute of the

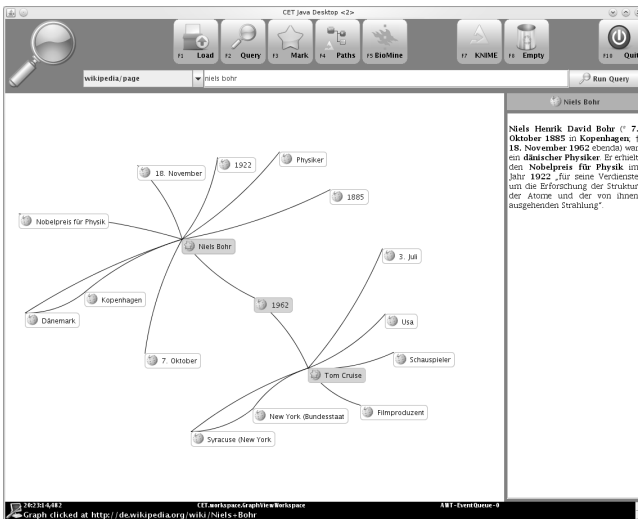


Fig. 1. Screenshot of the Creative Exploration Toolkit (CET)

nodes, while the actual calculation takes place in the underlying data analysis platform described in the next section. The user interface is therefore very flexible when it comes to tasks from different domains.

As described in the next section, any KNIME workflow may be called from inside the user interface. However, this is also meant for design and development purposes. Finalized workflows can be integrated more directly into the UI to provide a more natural and convenient user experience. In our current setup, we integrate calls to a shortest path calculation and queries to the BioMine² database. More workflow and interaction schemes will follow in the course of future work.

3 The KNIME Information Mining Platform

KNIME [1], the Konstanz Information Miner, was initially developed by the Chair for Bioinformatics and Information Mining at the University of Konstanz, Germany. KNIME is released under an open source license (GPL v3) and can be downloaded free of charge³. KNIME is a modular data exploration platform that enables the user to visually create data flows (often referred to as pipelines), selectively execute some or all analysis steps, and later investigate the results through interactive views on data and models. The KNIME base version already incorporates hundreds of processing nodes for data I/O, preprocessing and cleansing, modeling, analysis and data mining as well as various interactive views, such as scatter plots, parallel coordinates and others. It integrates all analysis modules of the well known Weka data mining environment and additional plugins allow, among others, R-scripts⁴ to be run, offering access to a vast library of statistical routines. Within the frame of the EU FP7 project “BISON”, KNIME was extended to also allow the flexible processing of large graphs. Combined with the already existing nodes, KNIME can therefore be used to model complex network processing and analysis tasks.

CET offers a very generic access to KNIME, enabling the user to make arbitrary calls without adapting the user interface. CET can be configured to directly call a KNIME workflow via a pre-configured button. CET also provides a list of all available workflows plus a list of parameters for a selected work, which can be edited by the user. Essentially, all information that would be sent by the user interface can be provided to start a KNIME workflow. The result is then visualized in the graph. New analysis methods can therefore be integrated easily into CET by simply adding a new workflow providing the corresponding functionality.

Figure 2 shows an example of a workflow computing the network diameter. In this workflow, firstly all nodes with a certain feature value are filtered, i.e. to take only those into account that have been selected and marked by the user. Secondly degree filters are applied on nodes and edges to filter unconnected

² <http://www.cs.helsinki.fi/group/biominer/>

³ <http://www.knime.org>

⁴ <http://www.r-project.org>

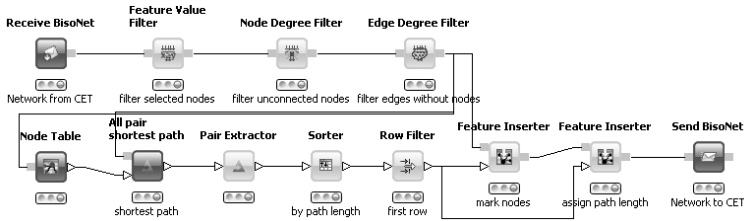


Fig. 2. An example KNIME workflow for calculating the network diameter which is called from CET

nodes. The shortest paths of all node pairs are subsequently computed and a feature is assigned consisting of the path length to those nodes of the longest of shortest paths. Finally the graph is sent back to the CET.

4 Conclusion and Future Work

We demonstrate a novel user interface for generic graph visualization with special emphasis on extensibility by integration with data and graph analysis. The presented interface allows for easy interaction with the visualized graphs. This setup is particularly interesting for researchers in the area of Data Mining and Network Analysis, as it is very simple to plug in new approaches and visualize the results, even if there is interaction involved.

Extensions of the CET aim towards the integration of more workflows, thus adding to the available interaction and analysis features. We will also further improve graph visualization by incorporating constraint-based graph layout (for a first discussion see [2]).

Acknowledgement. The work presented here was supported by the European Commission under the 7th Framework Programme FP7-ICT-2007-C FET-Open, contract no. BISON-211898.

References

1. Berthold, M.R., Cebron, N., Dill, F., Gabriel, T.R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., Wiswedel, B.: KNIME: The Konstanz Information Miner. In: Data Analysis, Machine Learning and Applications - Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Studies in Classification, Data Analysis, and Knowledge Organization, pp. 319–326. Springer, Heidelberg (2007)
2. Haun, S., Nitsche, M., Nürnberger, A.: Interactive Visualization of Continuous Node Features in Graphs. In: Proc. of Workshop on Explorative Analytics of Information Networks (EIN), part of ECML/PKDD 2009, pp. 98–106 (2009)
3. Koren, Y., Cıvırlı, A.: The Binary Stress Model for Graph Drawing. In: Tollis, I.G., Patrignani, M. (eds.) GD 2008. LNCS, vol. 5417, pp. 193–205. Springer, Heidelberg (2009)