MAVisto: a tool for the exploration of network motifs
Falk Schreiber and Henning Schwöbbermeyer
Leibniz Institute of Plant Genetics and Crop Plant Research, 06466 Gatersleben, Germany

ABSTRACT

Summary: MAVISTO is a tool for the exploration of motifs in biological networks. It provides a flexible motif search algorithm and different views for the analysis and visualisation of network motifs. These views help to explore interesting motifs: the frequency of motif occurrences can be compared to randomised networks, a list of motifs along with information about structure and number of occurrences depending on the reuse of network elements shows potentially interesting motifs, a motif fingerprint reveals the overall distribution of motifs of a given size, and the distribution of a particular motif in the network can be visualised using an advanced layout algorithm.

Availability: MAVISTO is platform independent and available free of charge as a Java webstart application at http://mavisto.ipk-gatersleben.de/.

Contact: schwoebb@ipk-gatersleben.de

Supplementary information: Can be found at http://mavisto.ipk-gatersleben.de/.

A general way to understand complex biological networks is to break them down into the simplest units of commonly used network architecture. Such patterns of local interconnection are called network motifs. They have been found in many different networks (Milo et al., 2002), but are particularly important for the understanding of biological networks. A well known motif is the feed-forward loop which performs key information processing roles in cells (Mangan & Alon, 2003). Further examples for the application of network motifs in biology are the prediction of interaction partners of proteins in protein-interaction networks (Albert & Albert, 2004), the classification of networks (Milo et al., 2004), and the analysis of structural network properties (Pržulj et al., 2004).

The analysis of network motifs is an important task in bioinformatics, however, it is not well supported by existing systems. To our knowledge only two tools for network motif analysis exist: MFINDER and PAJEK. The MFINDER network motif detection tool (Kashan et al., 2002) supports the numerical and statistical analysis of motifs in networks. It provides the results as text output, but without any means of visual analysis. On the other hand, PAJEK (Batagelj & Mrvar, 2004) is a graph analysis and visualisation tool which supports visual analysis tasks. However, it finds only motifs with three vertices (so called triangles) and therefore only a very restricted set of possible network motifs.

To support both the search for motifs of any size under different frequency concepts (that is different ways of counting motif occurrences depending on the reuse of network elements) and powerful exploration of motif distribution and motif fingerprint, we built a new tool called MAVISTO (Motif Analysis and VISualisation TOol). It is written in Java and based on Gravisto (Bachmaier et al., 2005), an editor for graphs and a toolkit for implementing graph algorithms. MAVISTO supports the Pajek-.net- (Batagelj & Mrvar, 2004) and the GML-format (Himsolt, 2000) for loading networks of interest and offers graph editor functionality for network manipulation and network creation from scratch, such as adding and moving vertices and edges. Furthermore, an advanced force-directed layout algorithm (Fruchterman & Reingold, 1991) is included to generate nice drawings of the network automatically which preserve the layout of motifs where possible.

MAVISTO's motif search algorithm discovers all motifs of a particular size. This size is either given by the number of vertices or by the number of edges. All motifs of this size are analysed and the frequencies for three different frequency concepts as well as the z-scores as a measure of the statistical significance are computed. The frequency concepts have different applications and restrictions on counting overlapping matches. Concept $F_1$ has no restrictions and considers all matches, therefore showing the full potential of a particular motif. Concept $F_2$ allows the sharing of vertices but not of edges and shows the number of instances of a motif which can be "active" at a time. For concept $F_3$ matches have to be vertex and edge disjoint and can be seen as non-overlapping clusters. The p-value and the z-score are obtained by comparing the frequency of all occurrences of a motif in the target network to the frequency values of this motif in an ensemble of randomisations of the target networks (Maslov et al., 2003). The algorithm for the search and the frequency concepts are described in detail in (Schreiber & Schwöbbermeyer, 2004).

MAVISTO facilitates the analysis of network motifs by presenting several views: (1) a list of motifs of the network along with information (called a motif table), (2) visual representations of motifs of interest (motif view), (3) a motif fingerprint (motif fingerprint) and (4) a visualisation of motif matches in the network (motif matches). (1) The motif table lists information such as the unique network motif label, the size of the motif, some structural properties and the different frequencies together with information about the statistical significance given by the p-value and the z-score. It allows sorting by all criteria and selecting of motifs to be displayed in the motif view. (2) The motif view provides a visual representation of the structure of motifs. (3) The motif fingerprint represents the motif frequency spectrum of the target network as a diagram. (4) The motif matches view provides the visual exploration of the occurrences of a motif within the analysed network and supports highlighting of the matches, respectively the covering of network elements by the matches, depending on the applied frequency concept. All different views are interwoven in a single interface. The views (1)-(3) allow selection of a motif and the active motif of other perspectives is updated accordingly. This coordination of different views significantly enhances the explorative power of network motif analysis.

*to whom correspondence should be addressed
Figure 1. Screen-shot of MAVISTO analysing a transcriptional regulatory network of Saccharomyces cerevisae with different perspectives to explore motifs. On the left side the network is shown with the motif preserving layout of highlighted matches of the feed-forward loop motif. On the right side all discovered motifs can be further analysed. Detailed information is presented in the motif table (top), the structure of the currently active motif is displayed in the motif view (middle) and the motif frequency spectrum is shown in the motif fingerprint (bottom).

Finding network motifs is computationally time consuming. As a general statement the search for motifs with four or five vertices in networks with 100-200 vertices and edges takes a few seconds, multiplied by the number of randomisations. The algorithm is accelerated for searching for motifs of size three to five in directed networks by the use of a lookup-table for motif isomorphism testing which maps graph labels based on the adjacency matrix to their canonical form.

ACKNOWLEDGEMENTS

This work was supported by the German Ministry of Education and Research (BMBF) under grant 0312706A. We would like to thank Christian Klukas for providing an implementation of the motif-preserving force-directed layout algorithm and the group of Franz J. Brandenburg (University of Passau, Germany) for kindly granting usage of Gravisto (http://www.gravisto.de/).

REFERENCES


