Application of network models to various biological problems

PhD thesis

Vilmos Ágoston

Supervisor:
Sándor Pongor PhD DSc

SZEGED
2007

Introduction

The success of network models is one of the most conspicuous scientific phenomena of the past few years. Importantly, network models can help us to recognize the common features of technical, computational as well as biological systems as they highlight the similar features in and the deep analogies between different fields. As a result they help to speed up sharing information between fields. For instance, network topologies, propagation of signals and material flows have many interesting common characteristics in vastly different fields, and it is widely believed that topological as well as flux-plasticity of network models may explain the robustness of biological systems.

The aim of my work was to get acquainted with the world of network models, and to apply them to various biological problems.

I chose three fields of application. The first of them was the examination of network stability or robustness. I have tested whether or not multiple but weak attacks

can be more efficient than single but strong attack. The second area was the mapping and representation of the network of oxidative folding intermediates of proteins. The third one was the characterization of protein similarity networks

Methods

The necessary programs were written in n C., using graph libraries such as Boost, Graphviz and Tulip. The programs were run on a 6 CPU Linux cluster.

In the examinations of network stability, I characterized robustness by the decrease of efficiency caused by random or targeted changes such as removal of edges or nodes. I compared various attack scenarios such as the removal of nodes, edges as well as changing the weight of the edges. In order to estimate the damage of multiple attacks I developed a greedy algorithm that chooses the maximum damage in successive steps.

Using simulated attacks I compared the effects of eliminating a node with the effects of halving the

Conferences:

Ágoston, V., P. Csermely, and S. Pongor. *Stability of biological networks and pharmacon design*, SZBK Straub-napok. 2004. Szeged.

Ágoston, V. *Multiple attacks on biological networks*, *GKEE Bioinformatics*. 2005. Balatonfüred.

Ágoston, V., P. Csermely, and S. Pongor. *Multiple weak hits confuse transcriptional regulatory networks*, FEBS. 2005. Budapest.

Pongor, S., Netotea, S. And Ágoston, V: *Modelling robust system topologies,* International Conference of Immunogenomics and Immunomics. 2006. Budapest.

Book chapters:

Ágoston, V., Kaján, L., Carugo, O., Hegedűs, Z., Vlahovicek, K. and Pongor, S. (2005) Concepts of similarity in bioinformatics In: Moss DS, Jelaska S and Pongor S., editors. Essays in Bioinformatics. IOS Press, Amsterdam, pp11-31.

Ágoston, V., Čemažar M., Pongor, S. (2005) Graph Representations of Oxidative Folding Pathways. In: Moss DS, Jelaska S and Pongor S., editors. Essays in Bioinformatics. IOS Press, Amsterdam, pp209-219

Kaján, L., Vlahovicek, K., Carugo, O., Ágoston, V., Hegedűs, Z. and Pongor, S. (2005) Comparison of sequences, protein 3D structures and genomes. In: Moss DS, Jelaska S and Pongor S., editors. Essays in Bioinformatics. IOS Press, Amsterdam, pp32-45.

number of a node's edges, weighting a node's edges and eliminating or weighting single edges. In all cases a greedy strategy was used in which algorithm selects the object whose removal causes the largest effect on the whole network.

Results

1. I examined the stability of biological (gene regulation) networks against multiple attacks. I worked out computational models for simulating the elimination or partial inactivation of network nodes (genes) and/or the contacts between them.
2. It was found that a few weak multiple perturbations can cause as much damage to a network as the elimination of a central object. This result draws attention to the possibilities of multitarget drug strategies as opposed to rationally designed, high specificity drugs.
3. I designed a network model to visualize the states corresponding to disulfide intermediates that can

form during the oxidative folding of proteins.

4. Experimentally studied intermediates can be mapped onto this network and the folding pathways appear as continuous paths connecting the reduced and the native state.

5. Similarity networks of protein sequences and structures were analyzed from a topological point of view. These networks have relatively high clustering coefficient characteristic of the called small world networks. In this kind of networks the similarity groups of proteins (such as domain types of protein families) appears as highly connected clusters.

6. The topologies were found database and method dependent. For instance, similarity networks built of the 3D structural similarities between members of protein domain data showed a hierarchical structure. On the other hand, the sequence similarity networks of the same data did not appear hierarchical.

7. The degree distribution of the analyzed similarity networks is reminiscent of the scale-free distribution, but not identical with it.

List of publication

Journal articles:

Ágoston, V., Csermely, P. and Pongor S. (2005) Weak hits confuse complex systems. *Physical Review E*, Stat Nonlin Soft Matter Phys. 2005 May; 71 (5Pt1):051909. Epub 2005 May 26.

Csermely, P. Ágoston, V., and Pongor S. (2005). The efficiency of multi-target drugs: The network approach may help drug design. *Trends in the Pharmacological Sciences*, **26**, 178-182.

Ágoston, V., Čemažar M., Kaján, L. and Pongor S. (2005) Graph representation of oxidative folding pathways. *BMC Bioinformatics*, **6**, 19.

Vlahovicek K., Kaján L, Ágoston V. and Pongor, S. (2005) The SBASE protein domain sequence resource, release 12.0: Prediction of Protein Domain Architecture using Support Vector Machines, Nucleic Acids Res., 33, D223-225.