

eXSys

Introduction

A feature of many complex networks is that only a relatively small proportion of individual network components are vital to its function, meaning that such systems are far more vulnerable to targeted attacks than to random component failures. Identifying these vulnerable components among a larger number of less important components can be computationally and analytically difficult, but is crucial for a large number of applications in a wide range of problem domains.

The aim of the eXSys project is to develop analysis tools to determine vulnerabilities in complex networks by identifying network components, connections and combinations of these, whose function and relative position in a network render them important to network integrity. Additionally, the project aims to develop novel methods of exploiting the knowledge generated from these tools, initially in the application domains of drug target discovery using protein interaction networks; the analysis of the effects of pesticides on the ecological systems where they are applied; and the design of secure information networks.

eXSys concentrates on the development of a software package to facilitate complex network analysis, providing support for the retrieval, integration and representation of the data required in order to model complex networks, the computational analysis of the modelled networks, and the presentation of results.

From network topology alone, statistical and graph theoretic methods can be used to analyse a network and determine important properties of both individual components and the network as a whole. The eXSys software will assist in the analysis of complex networks with the aim of developing techniques for exploiting the results obtained within three initial application domains:

- Drug target discovery using protein interaction networks, where the objective is to target proteins against which to aim pharmacological interventions. This process

involves identifying sets of proteins in a bacterium that allow selective destruction of the bacterium's metabolism, without adversely affecting the metabolism of a human host.

- The analysis of the effects of pesticides on ecological systems. This analysis can reveal the set of indirect effects caused by the application of pesticides, and can be used to predict the likely imbalances caused by the use of these compounds. Furthermore, network analysis can be used to design the efficient application of pesticide combinations such that their overall unbalancing effects on the whole ecosystem are minimized.
- The analysis and design of optimally failure- and attack-resistant information networks. Here, the topological properties of networks existing in large institutions will be investigated, where typically the risks of failure or attack are correspondingly large. Using network analysis techniques it should be possible to derive, for specific applications, schemes for optimal network connectivity and the distribution of processing and data. A chosen set of cost functions, such as cost of connectivity, robustness to failure of network elements or resistance to specifically targeted attacks, will be utilised to achieve this.

Data retrieval and integration

The information required to construct complex networks of this type usually exists in a variety of heterogeneous data sources. The eXSys project is initially focused on building data sets for intracellular metabolic/protein interaction networks, where protein interaction and metabolic interaction data may be obtained from databases such as the Database of Interacting Proteins (DIP) and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database. Additionally, interaction data may be gathered from annotations in protein sequence databases such as SwissProt, and conference/journal publications. Data retrieval involves automating the process of

extracting interaction data from the DIP, KEGG, and other databases, in addition to the development of algorithms for extracting information from textual data such as publication abstracts and sequence database annotations. This data is integrated into a common data format, and the data obtained combined to form a single data source consisting of protein interaction data for various species. The system automates the tasks of: downloading data from distributed databases; transforming data acquired into a common data format; integrating the data obtained into a single, comprehensive, consistent data set for as many organisms as possible; and keeping local data up to date and accurate by regularly examining remote databases. A range of Grid Services will be developed for the purpose of obtaining and integrating data for the various applications.

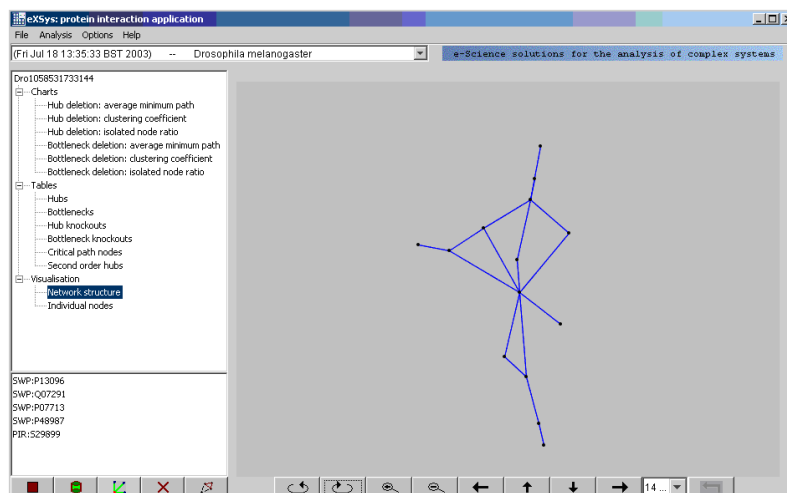
Network Analysis

A suite of network analysis algorithms are being developed to achieve the following objectives:

1. Identifying system components or connections whose removal or reduced function render the system structurally or functionally degraded
2. Identifying system components or connections whose removal or reduced function render a target system structurally or functionally degraded, while at the same time ensuring that the complex systems within which they are embedded are minimally affected
3. Identifying interventions that remediate the network when specific components are absent or damaged by restoring the network's structural integrity

User Interface

The eXSys software exposes a user interface allowing networks to be selected from a local database and analysed. Results are presented graphically using various tables, charts, and



graphs that highlight important properties of the network and its individual components. The interface integrates the various Grid Services.

Current and future research

Current research focuses on the development of the eXSys software and its application to the domains of drug target identification, the effects of pesticides on ecological networks, and the design of secure information networks. Many interesting networks are very large and their analysis is computationally demanding. Therefore eXSys will use Grid resources to gain access to high performance computing, enabling the analysis of larger networks than would otherwise be possible. Additionally, the project will investigate whether current Grid research in the area of data integration, such as OGSA-DAI and MyGrid can be harnessed to facilitate automatic data access and analysis across different data sources.

Investigators



Professor
Malcolm P. Young



Dr. Peter Andras

Research Associates



Dr. Olusola Idowu



Steven Lynden



Dr. Panayiotis
Periorellis

Robert Gwyther

Agnes Madalinski