

REVIEW

Open Access

# Construction and analysis of protein–protein interaction networks

Karthik Raman<sup>1,2\*</sup>

## Abstract

Protein–protein interactions form the basis for a vast majority of cellular events, including signal transduction and transcriptional regulation. It is now understood that the study of interactions between cellular macromolecules is fundamental to the understanding of biological systems. Interactions between proteins have been studied through a number of high-throughput experiments and have also been predicted through an array of computational methods that leverage the vast amount of sequence data generated in the last decade. In this review, I discuss some of the important computational methods for the prediction of functional linkages between proteins. I then give a brief overview of some of the databases and tools that are useful for a study of protein–protein interactions. I also present an introduction to network theory, followed by a discussion of the parameters commonly used in analysing networks, important network topologies, as well as methods to identify important network components, based on perturbations.

## Introduction

Proteins are the main catalysts, structural elements, signalling messengers and molecular machines of biological tissues [1]. Protein–protein interactions (PPIs) are extremely important in orchestrating the events in a cell. They form the basis for several signal transduction pathways in a cell, as well as various transcriptional regulatory networks. The availability of complete and annotated genome sequences of several organisms has led to a paradigm shift from the study of individual proteins in an organism to large-scale proteome-wide studies of proteins, which interact in a beautifully concerted network of metabolic, signalling and regulatory pathways in a cell. In general, the behaviour of a system is quite different from merely the sum of the interactions of its various parts. As Anderson put it as early as 1972, in his classic paper by the same title, “*More is different*” [2] — it is not possible to reliably predict the behaviour of a complex system, despite a good knowledge of the fundamental laws governing the individual components. Comparative genomics at a primary sequence level has also indicated that species differences are due more to the difference in the interactions between the component proteins, rather

than the individual genes themselves [3]. Consequently, several efforts have been made to identify these interactions, in an attempt to understand biological systems better [4-12]. The need to understand protein structure and function has been a critical driving force for biological research in the recent decades. With the advent of high-throughput experiments to identify PPIs, more knowledge on protein function has been obtained, together with the development of several methods to predict and study the interactions between proteins.

A wide variety of methods have been used to identify protein–protein associations; these associations may range from direct physical interactions inferred from experimental methods to functional linkages predicted on the basis of computational analyses. In the past, experimental methods based on microarrays and yeast two-hybrid, as well as computational methods based on protein sequences and structures have been developed and widely used. Given the difficulties in experimentally identifying PPIs, a wide range of computational methods have been used to identify protein–protein functional linkages and interactions. These methods range from identifying a single pair of interacting proteins at one end, to the identification and analysis of a large network of thousands of proteins, the latter as large as that of an entire proteome of a given cell.

\* Correspondence: [kraman@bioc.unizh.ch](mailto:kraman@bioc.unizh.ch)

<sup>1</sup>Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

## Computational methods for prediction of protein–protein functional linkages and interactions

### Methods based on genomic context

#### *Domain fusion*

The domain fusion or Rosetta Stone method was proposed by Eisenberg and co-workers [13]. The method is based on the hypothesis that if domains *A* and *B* exist fused in a single polypeptide *AB* in another organism, then *A* and *B* are functionally linked. Fig. 1A shows an example to illustrate this point. The premise is that since the affinity between proteins *A* and *B* is greatly enhanced when *A* is fused to *B*, some interacting pairs of proteins may have evolved from proteins that included the interacting domains *A* and *B* on the same polypeptide. Veitia [14] has proposed a kinetic background to the idea of gene fusion, suggesting the inclusion of eukaryotic sequences to increase the robustness of Rosetta Stone predictions. The argument basically involves the fact that eukaryotes, with a larger volume, cannot afford to accommodate separate proteins *A* and *B*, as the required concentrations of *A* and *B* would be prohibitively high, to achieve the same equilibrium concentration of *AB*. One limitation of this method is its low coverage; it has the least coverage among the methods based on genomic context [15].

#### *Conserved neighbourhood*

If the genes that encode two proteins are neighbours on the chromosome in several genomes, the corresponding proteins are likely to be functionally linked [16]. This method is particularly useful in case of prokaryotes, where operons commonly exist, or in organisms where operon-like clusters are observed. Fig. 1B shows an example to illustrate this method. This method has been reported to identify high-quality functional relationships [17]. However, the method suffers from low coverage, due to the dual requirement of identifying orthologues in another genome and then finding those orthologues that are adjacent on the chromosome [17]. Nevertheless, this coverage is still higher than that of the Rosetta Stone method [15]. Bork and co-workers have proposed another approach that exploits the conservation of divergently (bi-directionally) transcribed gene pairs [18]. The method is complementary to the existing gene neighbourhood method, which focuses on operons, where the genes are transcribed in a common orientation (co-directionally). They report the application of this method, to successfully associate self-regulatory transcription factors to their respective operons, enhancing functional annotations [18].

#### *Phylogenetic profiles*

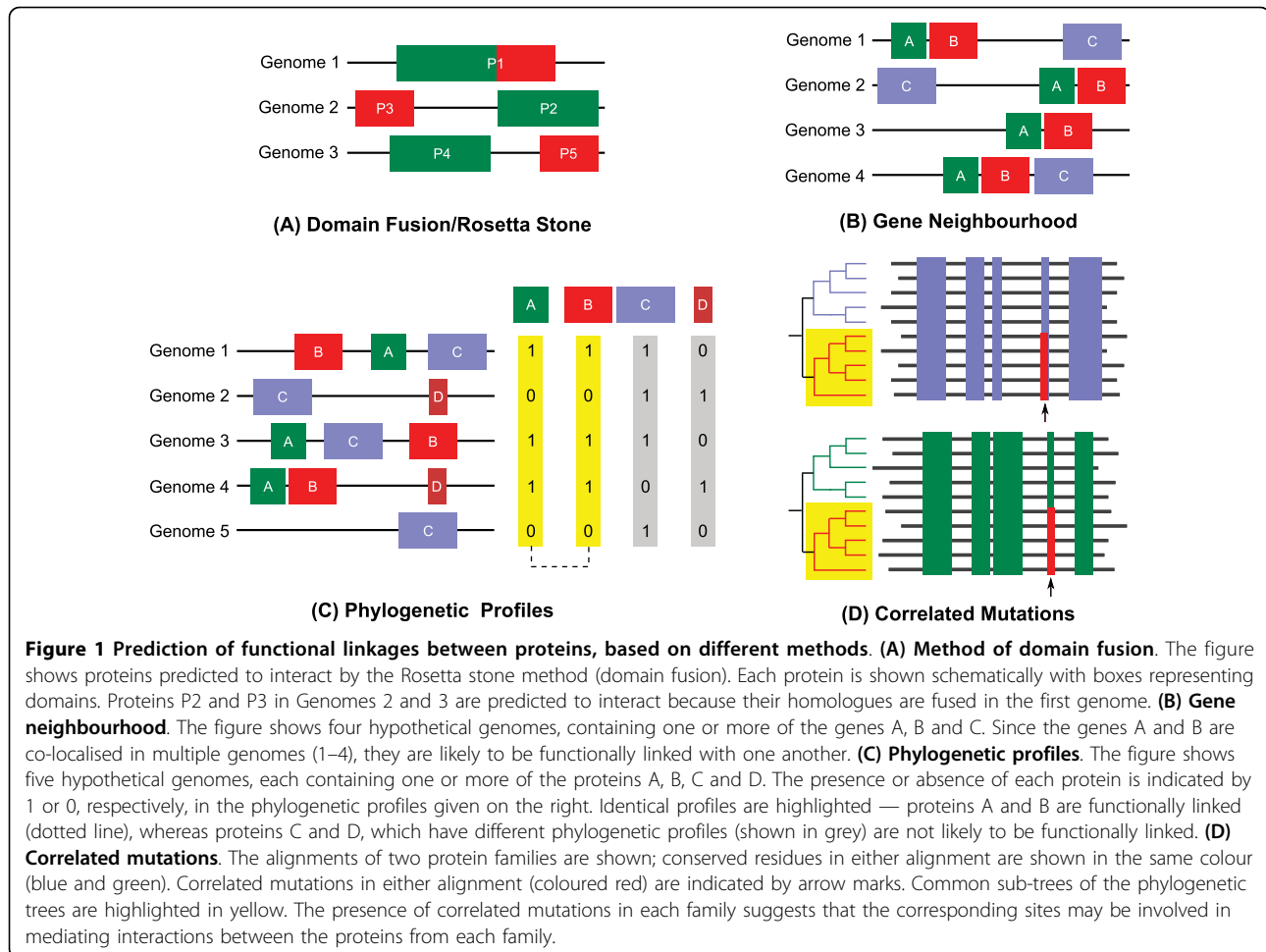
Identification of functional linkages between proteins using phylogenetic profiles is based on the idea that

functionally linked proteins would co-occur in genomes. The phylogenetic profile of a protein can be represented as a 'bit string', encoding the presence or absence of the protein in each of the genomes considered (see Fig. 1C). Proteins having matching or similar phylogenetic profiles tend to be strongly functionally linked [19]. In a study reported in 1999 [19], when only 17 fully sequenced genomes were considered for analysis, the function of a number of proteins in *Escherichia coli* could be assigned correctly, by examining the similarity of their phylogenetic profiles. Fig. 1C illustrates an example, showing how two proteins *A* and *B* are likely to be functionally linked, owing to the similarity of their phylogenetic profiles across five genomes. This method is in a sense the computational equivalent of the experimental genetic approach of mapping a mutant gene's phenotype to the gene. Genes with similar phylogenetic profiles essentially produce similar phenotypes, much similar to a standard genetic mapping [17]. Bork and co-workers [20] have used anti-correlated occurrences of genes (complementary phylogenetic patterns, as against co-occurrence) across genomes to identify several analogous enzyme displacements (functionally equivalent genes) in thiamine biosynthesis.

The online service Protein Link EXplorer (PLEX; <http://bioinformatics.icmb.utexas.edu/plex/>) [21] allows for the construction of phylogenetic profiles for any given sequence, which can be compared to profiles of all other proteins from 89 fully sequenced genomes that are stored in the PLEX database. PLEX can also accept sophisticated phylogenetic profile inputs and comparison parameters, including individual organism or group-based profiles. Gene neighbours and Rosetta stone links of all proteins that match the query profile can also be investigated.

### Methods based on co-evolution

Co-evolution can be defined as the joint evolution of ecologically interacting species [22] and it implies the evolution of a species in response to selection imposed by another. Co-evolution thus requires the existence of mutual selective pressure on two or more species [23]. Computational methods to predict PPIs through the characteristics of co-evolution have been developed by extrapolating concepts developed for the study of species co-evolution to the molecular level [23,24]. An *in silico* Two-hybrid (*i2h*) method has been proposed, based on the study of correlated mutations in multiple sequence alignments [25,26]. The premise is that co-adaptation of interacting proteins can be detected by the presence of a distinctive number of compensatory mutations in corresponding proteins of different species. An interaction index, defined based on the distribution of correlation values is calculated. Correlated mutations



can also be used to identify specific residues involved at the interaction sites [26]. Fig. 1D illustrates how correlated mutations can be used to identify functional linkages between proteins.

Protein interactions have also been predicted on the basis of the comparison of evolutionary histories, or phylogenetic trees, under the premise that interacting proteins are subject to similar evolutionary pressures resulting in similar topologies for the corresponding trees [27–29]. A more recent method [30] uses the complete network of phylogenetic tree similarities between all protein pairs in the genome to reassess pairwise similarity between the phylogenetic trees of any two proteins, thereby accounting for the co-evolutionary context of the proteins more effectively.

#### Other methods

Although homology-based methods are often quite useful for inferring PPIs, there are occasions where homology-based methods may not be effective. For example, Mika and Rost have illustrated earlier that homology-based inference of physical PPIs are accurate only at

high levels of sequence identity [31]. Further, homology-based inference of PPIs work better within species than across species, for low and high levels of sequence similarity [31].

Functional linkages may also be derived by the analysis of correlated mRNA expression levels, or protein co-expression. These techniques do not require any homology information [17], as they rely on the measurement of additional expression data. These techniques can, therefore, find unique relationships among proteins. The premise of all expression clustering methods is that proteins do not work in isolation and are often co-expressed with functionally related proteins. By altering the conditions for performing the experiments, enough variation in gene expression can be observed to identify co-expressing genes. Protein co-expression analysis is preferable since mRNA levels and protein levels have often been found to be poorly correlated.

Gene expression data has also been shown to be useful in understanding the dynamics of PPI networks [32–34]. Lu and collaborators [33] integrated gene

expression profiles (from a mice model of asthma) into a network of mouse PPIs derived from the BIND database. They found that highly connected proteins, or hub proteins in the network have less variable gene expression profiles compared to proteins at the network periphery. Mande and collaborators have described the construction of 'conditional networks' by integrating gene expression data under different conditions into protein functional linkage networks [34]. These networks present a picture of the dynamics of the functional linkages between proteins; a comparative analysis of four different conditional networks illustrates important responses in wild-type and mutant *Escherichia coli* cells treated with ultra-violet rays.

Efforts to mine experimental protein-protein association information from literature have also been made. For example, Hogue and co-workers have described an support vector machine (SVM)-based approach to mine the biomedical literature for PPIs [35]. Databases such as the STRING include such computationally mined interactions [36]. Eisenberg and co-workers have described an approach to identify abstracts that discuss PPIs from literature, which may then be manually scanned to identify PPIs [37]. This approach forms the basis for the rapid expansion of the database of interacting proteins (DIP) [37]. Zaki and collaborators have described a method based on pairwise similarity of protein sub-sequences, to predict PPIs [38].

### Experimental methods

Although this review primarily deals with computational methods for predicting PPIs, I here briefly outline some experimental methods for assessing PPIs, for the sake of completeness. There are a number of experimental techniques such as yeast-two hybrid [39], affinity purification/mass spectrometry [4,5,9,11,40] and protein microarrays [41-43], which are reviewed in detail elsewhere [44,45]. These form the basis of several large-scale datasets on PPIs.

In the yeast-two hybrid assay, two fusion proteins are created: the 'bait' (a protein of interest with a DNA-binding domain attached to its N-terminus) and the 'prey' (its potential interaction partner, fused to an activation domain). If the 'bait' and the 'prey' interact, their binding forms a functional transcriptional activator, which in turn activates reporter genes or selectable markers [39]. This assay has been adapted for high-throughput analyses of PPIs [46,47].

Gavin and collaborators have described the purification of complexes of 1739 proteins from *S. cerevisiae* (including the complete set of 1143 human orthologues) using tandem affinity purification coupled to mass spectrometry, illustrating the complexity of connectivity between protein complexes [4]. Mass spectrometry has

also been used to construct a large-scale map of human protein interactions [11].

Protein microarrays aid in the detection of *in vitro* binary interactions of various types — protein-protein, protein-lipid or antigen-antibody interactions. Proteins covalently attached to a solid support are screened with fluorescently labelled probes (proteins or lipids), to identify interactions [41]. A high density yeast protein microarray comprising 5800 yeast proteins was developed and used to identify novel calmodulin and phospholipid binding proteins [41].

Although many of these assays can identify PPIs with high confidence, they still have their share of false positives and can suffer from a limited reproducibility. Nevertheless, high-throughput experimental analyses of PPIs are quite important in obtaining the protein interaction map of a cell. Further, combining results from multiple experiments as well as computational methods for predicting functional linkages (as is done in databases such as the STRING) is likely to further improve our understanding of the complex web of interactions within a cell.

### Databases and tools for analysis of PPIs

In this section, I review some of the important databases that house data on PPIs, as well as some useful tools for the visualisation and analysis of PPIs. Protein interaction databases have also been reviewed in [44]. Some of the important databases containing data about PPIs are discussed below. Some more examples of databases useful for researching PPIs are given in Table 1.

#### STRING

STRING (Search Tool for the Retrieval of Interacting Genes/Proteins; <http://string.embl.de/>) [36,48] is a pre-computed database for the exploration and analysis of protein-protein associations. The associations are derived from high-throughput experimental data, mining of databases and literature, analyses of co-expressed genes and also from computational predictions, including those based on genomic context analysis. STRING employs a unique scoring framework based on benchmarks of the different types of associations against a common reference set, to produce a single confidence score per prediction. The graphical user interface is appealing and user-friendly, backed by an excellent visualisation engine. Medusa <http://coot.embl.de/medusa/>, a general graph visualisation tool, is a front end (interface) to the STRING protein interaction database [49].

#### HPRD

Human Protein Reference Database (HPRD; <http://www.hprd.org/>) [50] integrates information relevant to the

**Table 1 Databases and resources useful for researching PPIs.**

Database	URL	Resources	Refs.
BIND	Peer-reviewed bio-molecular interaction database containing published interactions and complexes	<a href="http://bind.ca/">http://bind.ca/</a>	[79]
BioGRID	Protein and genetic interactions from major model organism species	<a href="http://www.thebiogrid.org/">http://www.thebiogrid.org/</a>	[80]
COGs	Orthology data and phylogenetic profiles	<a href="http://www.ncbi.nlm.nih.gov/COG/">http://www.ncbi.nlm.nih.gov/COG/</a>	[81,82]
DIP	Experimentally determined interactions between proteins	<a href="http://dip.doe-mbi.ucla.edu/">http://dip.doe-mbi.ucla.edu/</a>	[51]
HPRD	Human protein functions, PPIs, post-translational modifications, enzyme–substrate relationships and disease associations	<a href="http://www.hprd.org/">http://www.hprd.org/</a>	[50,83]
IntAct	Interaction data abstracted from literature or from direct data depositions by expert curators	<a href="http://www.ebi.ac.uk/intact/">http://www.ebi.ac.uk/intact/</a>	[84]
iPFAM	Physical interactions between those Pfam domains that have a representative structure in the Protein DataBank (PDB)	<a href="http://ipfam.sanger.ac.uk/">http://ipfam.sanger.ac.uk/</a>	[85]
MINT	Experimentally verified PPI mined from the scientific literature by expert curators	<a href="http://mint.bio.uniroma2.it/mint/">http://mint.bio.uniroma2.it/mint/</a>	[86]
Predictome	Experimentally derived and computationally predicted functional linkages	<a href="http://visant.bu.edu/">http://visant.bu.edu/</a>	[52]
ProLinks	Protein functional linkages	<a href="http://mysql5.mbi.ucla.edu/cgi-bin/functionator/pronav">http://mysql5.mbi.ucla.edu/cgi-bin/functionator/pronav</a>	[87]
SCOPPI	Domain–domain interactions and their interfaces derived from PDB structure files and SCOP domain definitions	<a href="http://www.scoppi.org/">http://www.scoppi.org/</a>	[88]
STRING	Protein functional linkages from experimental data and computational predictions	<a href="http://string.embl.de/">http://string.embl.de/</a>	[36,48]

function of human proteins in health and disease. The database is almost completely manually curated by biologists who have read and interpreted over 300,000 published articles during the annotation process. Data pertaining to thousands of PPIs, post-translational modifications, enzyme/substrate relationships, disease associations, tissue expression and sub-cellular localisation have been extracted from literature into the database.

#### DIP

The DIP (Database of Interacting Proteins; <http://dip.doe-mbi.ucla.edu/>) database [51] catalogues experimentally derived PPIs. Due to the variety of experiments and their corresponding reliabilities, DIP applies some quality assessment methods to pick out subsets of most reliable interactions. The DIP is generally considered as a valuable benchmark or verify the performance of any new method for prediction of PPIs.

#### Predictome

The Predictome [52] database houses links between the proteins of 44 genomes based on the implementation of gene context functional linkage methods, viz. chromosomal proximity, phylogenetic profiling and domain fusion. It also contains information on large-scale experimental screenings of PPI data, from experiments such as yeast two-hybrid, immuno-co-precipitation and correlated expression. The Predictome database is presently accessible through the visual front-end provided by VisANT [53], which is a versatile tool for visualisation and analysis of interaction data. Website <http://visant.bu.edu/>.

#### Tools for network analysis and visualisation

In this section, I briefly discuss some of the useful software tools available for the analysis and visualisation of biological networks. A comprehensive review of the tools useful for the visualisation of networks has been published elsewhere [54]. Some more examples of tools useful for network visualisation and analysis are given in Table 2.

**Cytoscape** Cytoscape <http://www.cytoscape.org/>[55] is a software platform for visualising molecular interaction networks and integrating these interactions with gene expression profiles. The tool is best used in conjunction with large databases of gene expression data, protein–protein, protein–DNA, and genetic interactions that are increasingly available for humans and model organisms. Cytoscape supports several algorithms for the layout of networks. Several useful plug-ins are available for Cytoscape, to extend its capabilities. A notable example is the NetworkAnalyzer plug-in [56], which can be used to compute various network parameters.

**Pajek** Pajek <http://pajek.imfm.si/> is a program (only for Windows-based operating systems) for the analysis and visualisation of very large networks; it can even handle networks with  $> 10^5$  nodes. Pajek also includes a variety of network layout algorithms, including force-directed layout algorithms such as Fruchterman–Reingold [57]. Pajek is highly versatile and can also be used to study network dynamics.

#### Analyses of network structure

The field of network theory has witnessed a number of advances in the past [58–60], many of which are

**Table 2 Examples of tools useful for the visualisation of networks and PPIs.**

Tool	URL	Features	Refs.
BioLayout Express 3D	<a href="http://www.biolayout.org/">http://www.biolayout.org/</a>	Facilitates microarray data analysis	[89]
Cytoscape	<a href="http://www.cytoscape.org/">http://www.cytoscape.org/</a>	Versatile; implements many visualisation algorithms; many plug-ins available	[55]
Large Graph Layout (LGL)	<a href="http://sourceforge.net/projects/lgl">http://sourceforge.net/projects/lgl</a>	Especially useful for dynamic visualisation of large graphs ( $10^5$ nodes, $10^6$ edges); force-directed layout algorithm	[90]
Osprey	<a href="http://biodata.mshri.on.ca/osprey/servlet/Index">http://biodata.mshri.on.ca/osprey/servlet/Index</a>	Provides network filters, connectivity filters, many layouts and facilitates dataset superimposing	[91]
Pajek	<a href="http://vlado.fmf.uni-lj.si/pub/networks/pajek/">http://vlado.fmf.uni-lj.si/pub/networks/pajek/</a>	Especially useful for the analysis of very large networks	[92]
Visant	<a href="http://visant.bu.edu/">http://visant.bu.edu/</a>	Especially facilitates analysis of gene ontologies	[53]
Yed	<a href="http://www.yworks.com/products/yed/">http://www.yworks.com/products/yed/</a>	General purpose graph editor	-

impacting the analyses of biological networks such as PPI networks. In this section, I discuss some of the important network parameters useful in the analysis of networks and understanding their characteristics, important network topologies, as well as some of the measures that can be used to analyse perturbations to networks. Detailed reviews of the application of network theory to biology have been published elsewhere [61,62].

### Network parameters

Network theory provides a quantifiable description of networks; there are several network measures that enable the comparison and characterisation of complex networks:

#### Connectivity (or) Degree

The most elementary characteristic of a node is its degree,  $k$ , which represents the number of links the node has, to other nodes in the network.

#### Degree distribution

The degree distribution,  $P(k)$ , gives the probability that a selected node has exactly  $k$  links.  $P(k)$  is obtained by counting the number of nodes  $N(k)$  with  $k = 1, 2, \dots$  links and dividing by the number of nodes  $N$ . The degree distribution allows to distinguish between various network topologies [61].

#### Clustering Coefficient

The clustering coefficient was first defined by Watts and Strogatz [58]. The clustering coefficient,  $C$ , for a node is a notion of how connected the neighbours of a given node are (*cliquishness*). The average clustering coefficient for all nodes in a network is taken to be the network clustering coefficient. In an undirected graph, if a vertex  $v_i$  has  $k_i$  neighbours,  $k_i(k_i - 1)/2$  edges could exist among the vertices within the neighbourhood ( $N_i$ ). The clustering coefficient for an undirected graph  $\mathbf{G}(\mathbf{V}, \mathbf{E})$  (where  $\mathbf{V}$  represents the set of vertices in the graph  $\mathbf{G}$  and  $\mathbf{E}$  represents the set of edges) can then be defined as

$$C_i = \frac{2\{|e_{jk}\}|}{k_i(k_i-1)}; v_j, v_k \in N_i, e_{jk} \in \mathbf{E}. \quad (1)$$

The average clustering coefficient characterises the overall tendency of nodes to form clusters or groups.  $C(k)$  is defined as the average clustering coefficient for all nodes with  $k$  links.

#### Characteristic Path Length

The characteristic path length,  $L$ , is defined as the number of edges in the shortest path between two vertices, averaged over all pairs of vertices. It measures the typical separation between two vertices in the network [58]. Intuitively, it represents the network's overall navigability [61].

#### Network Diameter

The network diameter  $d$  is the greatest distance (shortest path, or geodesic path) between any two nodes in a network [63]. It can also be viewed as the length of the 'longest' *shortest path* in the network.

$$d = \max_{u, v \in \mathbf{G}} d_{\mathbf{G}}(u, v) \quad (2)$$

where  $d_{\mathbf{G}}(u, v)$  is the shortest path between  $u$  and  $v$  in  $\mathbf{G}$ . A few authors have also used this term to denote the average geodesic distance in a network (which translates to the characteristic path length), although strictly the two measures are distinct.

#### Betweenness

Betweenness is a centrality measure of a vertex within a graph [64]. For a graph  $\mathbf{G}(\mathbf{V}, \mathbf{E})$ , with  $n$  vertices, the betweenness  $C_B(v)$  of a vertex  $v$  is defined as

$$C_B(v) = \sum_{s \neq v \neq t \in \mathbf{V}} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (3)$$

where  $\sigma_{st}$  is the number of shortest paths from  $s$  to  $t$ , and  $\sigma_{st}(v)$  is the number of shortest paths from  $s$  to  $t$  that pass through a vertex  $v$ . A similar definition for

'edge betweenness' was given by Girvan and Newman [65]. Nodes with a higher betweenness lie on a larger number of shortest paths in a network.

### Network topologies

The understanding of the topology or the architectural principles of a biological network can directly give an insight into various network characteristics. There are several known topologies of networks, characterised by their distinctive network parameters. The following are some network models that are relevant to the understanding of biological networks.

#### Random networks

The Erdős–Rényi model of a random network starts with  $N$  nodes and connects each pair of nodes with a probability  $p$ , which creates a graph with approximately  $pN(N - 1)/2$  randomly placed links. The node degrees follow a Poisson distribution indicating that most nodes have approximately the same number of links. The characteristic path length is proportional to the logarithm of the network size  $L \sim \log N$ .  $C(k)$  is independent of  $k$  [61].

#### Small-world networks

Small-world networks are characterised by two properties: (i) individual nodes have few neighbours, but (ii) most nodes can be reached from one another through few steps, often referred to as 'six degrees of separation' [66]. Small-world networks have been generated by re-wiring regular ring-lattice-like networks [58]. A regular ring-lattice resembles a (circular) string of beads, where each node (bead) is linked to one node on either side, and is also additionally connected to the immediate neighbour of those nodes. Thus, each node is linked to four nodes nearest to it on the 'string'. The ring-lattice is rewired as follows: the original links in the lattice are replaced by random ones with a probability  $0 \leq \varphi \leq 1$ , introducing varying amounts of disorder, which takes the network from complete regularity to complete disorder (randomness). The re-wiring process allows the small-world model to interpolate between a regular lattice and a (more or less) random graph. When  $\varphi = 0$ , there is no re-wiring and the regular lattice remains unchanged. The clustering coefficient for this lattice tends to 0.75 for large  $k$ . The regular lattice, however, does not show the small-world effect. Mean geodesic distances between vertices tend to  $L/4k$  for large  $L$ . When  $\varphi = 1$ , every edge is re-wired to a new random location and the graph is almost a random graph, with typical geodesic distances on the order of  $\log L / \log k$ , but very low  $C \approx 2k/L$  [67]. As Watts and Strogatz showed by numerical simulation, however, there exists a sizeable region in between these two extremes of  $\varphi$ , for which the model generates a network that has both low path lengths and high clustering. Small-world networks

have a characteristic path length of the same order as random networks ( $L \propto \log N$ ), but have a clustering coefficient much higher than that of random networks ( $C \propto C_{\text{random}}$ ). The small-world topology has been observed in networks such as film actor networks, power grids and the neural network of the nematode *Caenorhabditis elegans* [58].

#### Scale-free Networks

Scale-free networks are characterised by a power-law degree distribution; the probability that a node has  $k$  links is given by  $P(k) \sim k^{-\gamma}$ , where  $\gamma$  is the degree exponent [59]. The value of  $\gamma$  determines many properties of the system. For smaller values of  $\gamma$ , the role of the 'hubs', or highly connected nodes, in the network becomes more important. For  $\gamma > 3$ , hubs are not relevant, while for  $2 < \gamma < 3$ , there is a hierarchy of hubs, with the most connected hub being in contact with a small fraction of all nodes. Scale-free networks have a high degree of robustness against random node failures, although they are sensitive to the failure of hubs. The probability that a node is highly connected is statistically more significant than in a random graph. The properties of a scale-free network are often determined by a relatively small number of highly connected hubs. The Barabási–Albert scale-free network model [59] involves the construction of a network through an iterative procedure. Beginning with a network having  $m_0$  nodes, in each subsequent iteration, a single node is added to the network, with  $m \leq m_0$  links to existing nodes. The probability with which this node connects to the existing nodes of the network is directly proportional to the connectivity of the existing nodes ('rich get richer' phenomenon). The probability  $p_i$  with which the new node connects to an existing node  $i$ , is given as

$$p_i = \frac{k_i}{\sum_{j \in \mathbf{G}} k_j}$$

where  $k_i$  is the degree of node  $i$  and the denominator represents the sum of the degrees of all nodes in the network ( $\mathbf{G}$ ). After  $n$  iterations, the model leads to a network with  $m_0 + n$  nodes and  $mn$  edges. The network generated by this model has a power-law degree distribution characterised by  $\gamma = 3$ . Scale-free networks with  $2 < \gamma < 3$ , a range commonly observed in many biological networks, are ultra-small, with a characteristic path length  $L \sim \log \log N$ , significantly smaller than that of random networks ( $\log N$ ) [61].

#### Analysis of network perturbations

Networks can be perturbed through the removal of nodes and edges. A typical analysis would be to probe the effect of disrupting a node and its corresponding edges. Networks of different topologies vary in their resilience to

various types of perturbations. A number of studies have been carried out to analyse the response of networks to the deletion of their nodes and edges. A review of how nodes in a network can be prioritised based on network analysis has been presented elsewhere [68].

Barabási and co-workers have analysed the response of scale-free and random networks to various types of 'attacks' [69]. In particular, they have analysed the networks representing the topologies of the Internet and the World-Wide Web. The common observation is that scale-free networks are quite insensitive to random node removals; they are highly robust in the face of random node failures and the characteristic path length was found to be almost unaffected. This is intuitively reasonable, since most of the vertices in these networks have low degree and therefore lie on few paths between others; thus their removal rarely affects communications substantially. On the other hand, directed attacks targeting the highly connected hubs led to a rapid disruption of the communication through the network. The characteristic path length was found to increase very sharply with the fraction of hubs removed and typically only a small fraction of the hubs needed to be 'knocked out' before essentially all communication through the network was destroyed [67,69].

Jeong and co-workers have analysed the effect of node deletions on *S. cerevisiae* PPI network [70]. They report that although proteins with five or fewer links constituted about 93% of the total number of proteins, only about 21% of them were essential. On the other hand, only 0.7% of the proteins had more than 15 links, but single deletion of 62% of these proved lethal. This implies that highly connected proteins with a central role in the architecture of the network are three times more likely to be essential than proteins with only a small number of links to other proteins.

Another comprehensive analysis of vulnerability of complex networks to various types of attacks has been discussed in [71]. In addition to node deletions studied earlier [69], they have also studied the effects of edge removals. Further, for each case of attacks on vertices and edges, four different attacking strategies were employed: removals by the descending order of the degree and the betweenness centrality, calculated for either the initial network or the modified network during the iterative removal procedure. They report that the removals based on the re-calculated degrees and betweenness centralities are often more harmful than the attack strategies based on the initial network's parameters, underlining the importance of the changes in network structure following the removal of important edges or nodes.

Wingender and co-workers have proposed a measure, known as *pairwise disconnectivity index* [72], which

quantifies how crucial a node or an edge (or a group of nodes/edges) is, for sustaining the communication between connected pairs of vertices in a directed network. This is one metric that explicitly considers paths between the various nodes in a network; it is thus quite useful in analysing how node deletions in a network can disrupt the flow of information.

We have earlier reported an analysis of the number of disrupted shortest paths in the network, to identify nodes that may be critical to a network [73]. Network analysis has also been used for identifying pathways to drug resistance [74]. Ge and collaborators have developed an 'information flow analysis', to identify proteins central for information transmission in interactome networks of *S. cerevisiae* and *C. elegans* [75]; the proteins so identified were also likely to be essential for survival. The method employs confidence scores for PPIs and also considers multiple paths in a network while evaluating the importance of each protein [75]. The analysis of node deletions from PPI networks has been used for the identification of potential drug targets [73,76].

## Conclusions

PPI networks provide a simplified overview of the web of interactions that take place inside a cell. The vast amounts of sequence data that have been generated have been leveraged to make better predictions of interactions and functional associations between proteins, as well as individual protein functions. By integrating experimental methods for determining PPIs and computational methods for prediction, a lot of useful data on PPIs have been generated, including a number of high-quality databases.

Although the analyses of PPI networks has produced several useful results, often improving our understanding of the underlying biology, they are not without flaws. One of the key flaws of the existing methods to delineate such large-scale protein interaction networks is the limited reproducibility of such experiments; further, it is suspected that what is examined is only a small fraction of the entire proteome [77]. However, most databases do combine multiple methods for predicting interactions, as well as results from multiple high-throughput experiments, mitigating this problem to a certain extent. Further, these networks often paint a static picture of the overwhelmingly complex dynamic interactions that take place in a cell. An improved model of these interactions must consider both the dynamics (temporal changes in the interactions) as well as the strengths of each of the interactions. The global overview presented by such interaction maps is no doubt useful, but the finer details of the interactions may be significantly important for our ability to make testable predictions about biological systems [78].



Nevertheless, protein interaction maps have many practical applications and hold the key to understanding complex biological systems. With a large amount of high-throughput data being generated at various levels, computational analyses of these data, to identify associations and interactions between various proteins, form a fundamental step in our quest to understand the organisation of complex biological systems. As Dennis Bray put it rather eloquently [78], “*We have a new continent to explore and will need maps at every scale to find our way*”.

#### Acknowledgements

The author is grateful to Nagasuma Chandra and Andreas Wagner for their mentorship. Financial support through the YeastX project of SystemsX.ch is gratefully acknowledged.

#### Author details

<sup>1</sup>Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland. <sup>2</sup>Swiss Institute of Bioinformatics, Quartier Sorge, Batiment Genopode, 1015 Lausanne, Switzerland.

#### Authors' contributions

KR wrote, read and approved the final manuscript.

#### Competing interests

The author declares that he has no competing interests.

Received: 25 November 2009

Accepted: 15 February 2010 Published: 15 February 2010

#### References

1. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO: **Protein function in the post-genomic era.** *Nature* 2000, **405(6788)**:823-826.
2. Anderson PW: **More Is Different.** *Science* 1972, **177(4047)**:393-396.
3. Valencia A, Pazos F: **Computational methods for the prediction of protein interactions.** *Curr Opin Struct Biol* 2002, **12**:368-373.
4. Gavin AC, Bäsch M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hfert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G: **Functional organization of the yeast proteome by systematic analysis of protein complexes.** *Nature* 2002, **415(6868)**:141-147.
5. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, Yang L, Wolting C, Donaldson I, Schandorff S, Shewnarane J, Vo M, Taggart J, Goudreau M, Muskut B, Alfarano C, Dewar D, Lin Z, Michalickova K, Willems AR, Sassi H, Nielsen PA, Rasmussen KJ, Andersen JR, Johansen LE, Hansen LH, Jespersen H, Podtelejnikov A, Nielsen E, Crawford J, Poulsen V, Srensen BD, Matthiesen J, Hendrickson RC, Gleeson F, Pawson T, Moran MF, Durocher D, Mann M, Hogue CW, Figeys D, Tyers M: **Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry.** *Nature* 2002, **415(6868)**:180-183.
6. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E, Vijayadamar G, Pochart P, Machineni H, Welsh M, Kong Y, Zerhusen B, Malcolm R, Varrone Z, Collis A, Minto M, Burgess S, McDaniel L, Stimpson E, Spriggs F, Williams J, Neurath K, Ioime N, Agee M, Voss E, Furtak K, Renzulli R, Aanensen N, Carrolla S, Bickelhaupt E, Lazovatsky Y, DaSilva A, Zhong J, Stanyon CA, Finley RL, White KP, Braverman M, Jarvie T, Gold S, Leach M, Knight J, Shimkets RA, McKenna MP, Chant J, Rothberg JM: **A protein interaction map of *Drosophila melanogaster*.** *Science* 2003, **302(5651)**:1727-1736.
7. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Heuvel SVD, Piano F, Vandenhaute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M: **A map of the interactome network of the metazoan *C. elegans*.** *Science* 2004, **303(5657)**:540-543.
8. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamosas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M: **Towards a proteome-scale map of the human protein-protein interaction network.** *Nature* 2005, **437(7062)**:1173-1178.
9. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, Datta N, Tikuisis AP, Punna T, Peregrn-Alvarez JM, Shales M, Zhang X, Davey M, Robinson MD, Paccanaro A, Bray JE, Sheung A, Beattie B, Richards DP, Canadien V, Lalev A, Mena F, Wong P, Starostine A, Canete MM, Vlasblom J, Wu S, Orsi C, Collins SR, Chandran S, Haw R, Rilstone JJ, Gandi K, Thompson NJ, Musso G, Ong PS, Ghanny S, Lam MHY, Butland G, Altaf-Ul AM, Kanaya S, Shilatifard A, O'Shea E, Weissman JS, Ingles CJ, Hughes TR, Parkinson J, Gerstein M, Wodak SJ, Emili A, Greenblatt JF: **Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*.** *Nature* 2006, **440(7084)**:637-643.
10. Arifuzzaman M, Maeda M, Itoh A, Nishikata K, Takita C, Saito R, Ara T, Nakahigashi K, Huang HC, Hirai A, Tsuzuki K, Nakamura S, Altaf-Ul-Amin M, Oshima T, Baba T, Yamamoto N, Kawamura T, Ioka-Nakamichi T, Kitagawa M, Tomita M, Kanaya S, Wada C, Mori H: **Large-scale identification of protein-protein interaction of *Escherichia coli* K-12.** *Genome Res* 2006, **16(5)**:686-691.
11. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M, Taylor R, Dharsee M, Ho Y, Heilbut A, Moore L, Zhang S, Ornatsky O, Bukhman YV, Ethier M, Sheng Y, Vasilescu J, Abu-Farha M, Lambert JP, Duewel HS, Stewart II, Kuehl B, Hogue K, Colwill K, Gladwish K, Muskut B, Kinach R, Adams SL, Moran MF, Morin GB, Topaloglou T, Figeys D: **Large-scale mapping of human protein-protein interactions by mass spectrometry.** *Mol Syst Biol* 2007, **3**:89.
12. Yu H, Braun P, Yildirim MA, Lemmens I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N, Hao T, Rual JF, Dricot A, Vazquez A, Murray RR, Simon C, Tardivo L, Tam S, Svrikapa N, Fan C, de Smet AS, Motyl A, Hudson ME, Park J, Xin X, Cusick ME, Moore T, Boone C, Snyder M, Roth FP, Barabasi AL, Tavernier J, Hill DE, Vidal M: **High-quality binary protein interaction map of the yeast interactome network.** *Science* 2008, **322(5898)**:104-110.
13. Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO, Eisenberg D: **Detecting Protein Function and Protein-Protein Interactions from Genome Sequences.** *Science* 1999, **285(5428)**:751-753.
14. Veitia RA: **Rosetta Stone proteins: "chance and necessity"?** *Genome Biol* 2002, **3(2)**:interactions1001.1-1001.3.
15. Huynen M, Snel B, Lathe W, Bork P: **Predicting protein function by genomic context: quantitative evaluation and qualitative inferences.** *Genome Res* 2000, **10(8)**:1204-1210.
16. Dandekar T, Snel B, Huynen MA, Bork P: **Conservation of gene order: a fingerprint of proteins that physically interact.** *Trends Biochemical Sci* 1998, **23(9)**:324-328.
17. Marcotte EM: **Computational genetics: finding protein function by nonhomology methods.** *Curr Opin Struct Biol* 2000, **10**:359-365.
18. Korbel JO, Jensen LJ, von Mering C, Bork P: **Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs.** *Nat Biotechnol* 2004, **7**:911-917.
19. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO: **Assigning protein functions by comparative genome analysis: Protein phylogenetic profiles.** *Proc Natl Acad Sci USA* 1999, **96(8)**:4285-4288.
20. Morett E, Korbel JO, Rajan E, Saab-Rincon G, Olvera L, Olvera M, Schmidt S, Snel B, Bork P: **Systematic discovery of analogous enzymes in thiamin biosynthesis.** *Nat Biotechnol* 2003, **21**:790-795.
21. Date SV, Marcotte EM: **Protein function prediction using the Protein Link Explorer (PLEX).** *Bioinformatics* 2005, **21(10)**:2558-2559.
22. Thompson J: *The Coevolutionary Process* Chicago: University of Chicago Press 1994.

23. Pazos F, Valencia A: **Protein co-evolution, co-adaptation and interactions.** *EMBO J* 2008, **27(20)**:2648-2655.
24. Barker D, Pagel M: **Predicting functional gene links from phylogenetic-statistical analyses of whole genomes.** *PLoS Comput Biol* 2005, **1**:e3.
25. Pazos F, Helmer-Citterich M, Ausiello G, Valencia A: **Correlated mutations contain information about protein-protein interaction.** *J Mol Biol* 1997, **271(4)**:511-523.
26. Pazos F, Valencia A: **In silico Two-Hybrid System for the Selection of Physically Interacting Protein Pairs.** *Proteins* 2002, **47**:219-227.
27. Goh CS, Cohen FE: **Co-evolutionary analysis reveals insights into protein-protein interactions.** *J Mol Biol* 2002, **324**:177-192.
28. Ramani AK, Marcotte EM: **Exploiting the co-evolution of interacting proteins to discover interaction specificity.** *J Mol Biol* 2003, **327**:273-284.
29. Pazos F, Ranea JAG, Juan D, Sternberg MJ: **Assessing protein co-evolution in the context of the tree of life assists in the prediction of the interactome.** *J Mol Biol* 2005, **352(4)**:1002-1015.
30. Juan D, Pazos F, Valencia A: **High-confidence prediction of global interactomes based on genome-wide coevolutionary networks.** *Proc Natl Acad Sci USA* 2008, **105(3)**:934-939.
31. Mika S, Rost B: **Protein-protein interactions more conserved within species than across species.** *PLoS Comput Biol* 2006, **2(7)**:e79.
32. Komurov K, White M: **Revealing static and dynamic modular architecture of the eukaryotic protein interaction network.** *Mol Syst Biol* 2007, **3**:110.
33. Lu X, Jain W, Finn PW, Perkins DL: **Hubs in biological interaction networks exhibit low changes in expression in experimental asthma.** *Mol Syst Biol* 2007, **3**:98.
34. Hegde SR, Manimaran P, Mande SC: **Dynamic changes in protein functional linkage networks revealed by integration with gene expression data.** *PLoS Comput Biol* 2008, **4(11)**:e1000237.
35. Donaldson I, Martin J, de Bruijn B, Wolting C, Lay V, Tuekam B, Zhang S, Baskin B, Bader GD, Michalickova K, Pawson T, Hogue CWV: **PreBIND and Textomy - mining the biomedical literature for protein-protein interactions using a support vector machine.** *BMC Bioinformatics* 2003, **4**:11.
36. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P: **STRING: known and predicted protein-protein associations, integrated and transferred across organisms.** *Nucleic Acids Res* 2005, **33(Suppl 1)**:D433-437.
37. Marcotte EM, Xenarios I, Eisenberg D: **Mining literature for protein-protein interactions.** *Bioinformatics* 2001, **17(4)**:359-363.
38. Zaki N, Lazarova-Molnar S, El-Hajj W, Campbell P: **Protein-protein interaction based on pairwise similarity.** *BMC Bioinformatics* 2009, **10**:150.
39. Fields S, Song O: **A novel genetic system to detect protein-protein interactions.** *Nature* 1989, **340(6230)**:245-246.
40. Gingras AC, Gstaiger M, Raught B, Aebersold R: **Analysis of protein complexes using mass spectrometry.** *Nat Rev Mol Cell Biol* 2007, **8(8)**:645-654.
41. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, Mitchell T, Miller P, Dean RA, Gerstein M, Snyder M: **Global analysis of protein activities using proteome chips.** *Science* 2001, **293(5537)**:2101-2105.
42. Michaud GA, Salcius M, Zhou F, Bangham R, Bonin J, Guo H, Snyder M, Predki PF, Schweitzer BI: **Analyzing antibody specificity with whole proteome microarrays.** *Nat Biotechnol* 2003, **21(12)**:1509-1512.
43. Mattoon DR, Schweitzer B: **Profiling protein interaction networks with functional protein microarrays.** *Methods Mol Biol* 2009, **563**:63-74.
44. Shoemaker BA, Panchenko AR: **Deciphering protein-protein interactions. Part I. Experimental techniques and databases.** *PLoS Comput Biol* 2007, **3(3)**:e42.
45. Uetz P: **Experimental methods for protein interaction identification and characterization.** *Protein-protein interactions and networks* Springer 2008:1-32.
46. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamar G, Yang M, Johnston M, Fields S, Rothberg JM: **A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*.** *Nature* 2000, **403(6770)**:623-627.
47. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y: **A comprehensive two-hybrid analysis to explore the yeast protein interactome.** *Proc Natl Acad Sci USA* 2001, **98(8)**:4569-4574.
48. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C: **STRING 8-a global view on proteins and their functional interactions in 630 organisms.** *Nucleic Acids Res* 2009, **37 Database**: D412-D416.
49. Hooper SD, Bork P: **Medusa: a simple tool for interaction graph analysis.** *Bioinformatics* 2005, **21(24)**:4432-4433.
50. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjan V, Muthusamy B, Gandhi TKB, Gronborg M, Ibarrola N, Deshpande N, Shanker K, Shivashankar HN, Rashmi BP, Ramya MA, Zhao Z, Chandrika KN, Padma N, Harsha HC, Yatish AJ, Kavitha MP, Menezes M, Choudhury DR, Suresh S, Ghosh N, Saravana R, Schiemann WP, Constantinescu SN, Huang L, Khosravi-Far R, Steen H, Tewari M, Ghaffari S, Blobel GC, Dang CV, Garcia JGN, Pevsner J, Jensen ON, Roepstorff P, Deshpande KS, Chinnaiyan AM, Hamosh A, Chakravarti A, Pandey A: **Development of human protein reference database as an initial platform for approaching systems biology in humans.** *Genome Res* 2003, **13(10)**:2363-2371.
51. Xenarios I, Fernandez E, Salwinski L, Duan XJ, Thompson MJ, Marcotte EM, Eisenberg D: **DIP: The Database of Interacting Proteins: 2001 update.** *Nucleic Acids Res* 2001, **29**:239-241.
52. Mellor JC, Yanai I, Clodfelter KH, Mintseris J, DeLisi C: **Predictome: a database of putative functional links between proteins.** *Nucleic Acids Res* 2002, **30**:306-309.
53. Hu Z, Snitkin ES, DeLisi C: **VisANT: an integrative framework for networks in systems biology.** *Brief Bioinform* 2008, **9(4)**:317-325.
54. Pavlopoulos G, Wegener AL, Schneider R: **A survey of visualization tools for biological network analysis.** *BioData Min* 2008, **1**:12.
55. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13(11)**:2498-2504.
56. Assenov Y, Ramirez F, Schelhorn SE, Lengauer T, Albrecht M: **Computing topological parameters of biological networks.** *Bioinformatics* 2008, **24(2)**:282-284.
57. Fruchterman TMJ, Reingold EM: **Graph drawing by force-directed placement.** *Softw Pract Exper* 1991, **21(11)**:1129-1164.
58. Watts DJ, Strogatz SH: **Collective dynamics of 'small-world' networks.** *Nature* 1998, **393(6684)**:440-442.
59. Barabási AL, Albert R: **Emergence of Scaling in Random Networks.** *Science* 1999, **286(5439)**:509-512.
60. Albert R, Jeong H, Barabási AL: **Diameter of the World-Wide Web.** *Nature* 1999, **401**:130-131.
61. Barabási AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nat Rev Genet* 2004, **5(2)**:101-113.
62. Mason O, Verwoerd M: **Graph theory and networks in Biology.** *IET Syst Biol* 2007, **1(2)**:89-119.
63. Diestel R: **Graph Theory.** *Graduate Texts in Mathematics* Springer-Verlag 2000, **173**.
64. Freeman LC: **A set of measures of centrality based on betweenness.** *Sociometry* 1977, **40**:35-41.
65. Girvan M, Newman MEJ: **Community structure in social and biological networks.** *Proc Natl Acad Sci USA* 2002, **99(12)**:7821-7826.
66. Watts D: *Six Degrees* London: W. W. Norton & Company 2003.
67. Newman MEJ: **The Structure and Function of Complex Networks.** *SIAM Review* 2003, **45(2)**:167-256.
68. Chang AN: **Prioritizing genes for pathway impact using network analysis.** *Methods Mol Biol* 2009, **563**:141-156.
69. Albert R, Jeong H, Barabási AL: **Error and attack tolerance of complex networks.** *Nature* 2000, **406(6794)**:378-382.
70. Jeong H, Mason SP, Barabási AL, Oltvai ZN: **Lethality and centrality in protein networks.** *Nature* 2001, **411(6833)**:41-42.
71. Holme P, Kim BJ, Yoon CN, Han SK: **Attack vulnerability of complex networks.** *Phys Rev E* 2002, **65(5)**:056109.
72. Potapov AP, Goemann B, Wingender E: **The pairwise disconnectivity index as a new metric for the topological analysis of regulatory networks.** *BMC Bioinformatics* 2008, **9**:227.
73. Raman K, Kalidas Y, Chandra N: **targetTB: A target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis.** *BMC Syst Biol* 2008, **2**:109.

74. Raman K, Chandra N: ***Mycobacterium tuberculosis* interactome analysis unravels potential pathways to drug resistance.** *BMC Microbiol* 2008, **8**:234.
75. Missiuro PV, Liu K, Zou L, Ross BC, Zhao G, Liu JS, Ge H: **Information flow analysis of interactome networks.** *PLoS Comput Biol* 2009, **5**(4):e1000350.
76. Raman K, Vashisht R, Chandra N: **Strategies for efficient disruption of metabolism in *Mycobacterium tuberculosis* from network analysis.** *Mol Biosyst* 2009, **5**:1740-1751.
77. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P: **Comparative assessment of large-scale data sets of protein-protein interactions.** *Nature* 2002, **417**(6887):399-403.
78. Bray D: **Molecular networks: the top-down view.** *Science* 2003, **301**(5641):1864-1865.
79. Bader GD, Betel D, Hogue CW: **BIND: the Biomolecular Interaction Network Database.** *Nucleic Acids Res* 2003, **31**:248-250.
80. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M: **BioGRID: a general repository for interaction datasets.** *Nucleic Acids Res* 2006, **34** Database: D535-D539.
81. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278**(5338):631-637.
82. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA: **The COG database: an updated version includes eukaryotes.** *BMC Bioinformatics* 2003, **4**:41.
83. Prasad TSK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A, Balakrishnan L, Marimuthu A, Banerjee S, Somanathan DS, Sebastian A, Rani S, Ray S, Kishore CJH, Kanth S, Ahmed M, Kashyap MK, Mohmood R, Ramachandra YL, Krishna V, Rahiman BA, Mohan S, Ranganathan P, Ramabadran S, Chaerkady R, Pandey A: **Human Protein Reference Database-2009 update.** *Nucleic Acids Res* 2009, **37** Database: D767-D772.
84. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, Feuermann M, Ghanbarian AT, Kerrien S, Khadake J, Kerssemakers J, Leroy C, Menden M, Michaut M, Montecchi-Palazzi L, Neuhauser SN, Orchard S, Perreau V, Roechert B, van Eijk K, Hermjakob H: **The IntAct molecular interaction database in 2010.** *Nucleic Acids Res* 2009.
85. Finn RD, Marshall M, Bateman A: **iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions.** *Bioinformatics* 2005, **21**(3):410-412.
86. Chatr-aryamontri A, Ceol A, Palazzi LM, Nardelli G, Schneider MV, Castagnoli L, Cesareni G: **MINT: the Molecular INTERaction database.** *Nucleic Acids Res* 2007, **35** Database: D572-D574.
87. Bowers PM, Pellegrini M, Thompson MJ, Fierro J, Yeates TO, Eisenberg D: **Prolinks: a database of protein functional linkages derived from coevolution.** *Genome Biol* 2004, **5**:R35.
88. Winter C, Henschel A, Kim WK, Schroeder M: **SCOPPI: a structural classification of protein-protein interfaces.** *Nucleic Acids Res* 2006, **34** Database: D310-D314.
89. Freeman TC, Goldovsky L, Brosch M, van Dongen S, Mazire P, Grocock RJ, Freilich S, Thornton J, Enright AJ: **Construction, visualisation, and clustering of transcription networks from microarray expression data.** *PLoS Comput Biol* 2007, **3**(10):2032-2042.
90. Adai AT, Date SV, Wieland S, Marcotte EM: **LGL: creating a map of protein function with an algorithm for visualizing very large biological networks.** *J Mol Biol* 2004, **340**:179-190.
91. Breitkreutz BJ, Stark C, Tyers M: **Osprey: a network visualization system.** *Genome Biol* 2003, **4**(3):R22.
92. Batagelj V, Mrvar A: **Pajek - Program for Large Network Analysis.** *Connections* 1998, **21**:47-57 <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=>

doi:10.1186/1759-4499-2-2

**Cite this article as:** Raman: Construction and analysis of protein-protein interaction networks. *Automated Experimentation* 2010 2:2.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

