DOI 10.1002/pmic.200700095 *Proteomics* 2007, *7*, 2856–2869

REVIEW

Discovering functions and revealing mechanisms at molecular level from biological networks

*Shihua Zhang[1], Guangxu Jin[1], Xiang-Sun Zhang[1] and Luonan Chen[2, 3, 4, 5]**

[1] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
[2] Institute of Systems Biology, Shanghai University, Shanghai, China
[3] Osaka Sangyo University, Osaka, Japan
[4] ERATO Aihara Complexity Modelling Project, JST, Tokyo, Japan
[5] Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

Received: February 2, 2007
Revised: April 12, 2007
Accepted: April 12, 2007

With the increasingly accumulated data from high-throughput technologies, study on biomolecular networks has become one of key focuses in systems biology and bioinformatics. In particular, various types of molecular networks (*e.g.*, protein–protein interaction (PPI) network; gene regulatory network (GRN); metabolic network (MN); gene coexpression network (GCEN)) have been extensively investigated, and those studies demonstrate great potentials to discover basic functions and to reveal essential mechanisms for various biological phenomena, by understanding biological systems not at individual component level but at a system-wide level. Recent studies on networks have created very prolific researches on many aspects of living organisms. In this paper, we aim to review the recent developments on topics related to molecular networks in a comprehensive manner, with the special emphasis on the computational aspect. The contents of the survey cover global topological properties and local structural characteristics, network motifs, network comparison and query, detection of functional modules and network motifs, function prediction from network analysis, inferring molecular networks from biological data as well as representative databases and software tools.

Keywords:
Biological network / Network analysis / Network comparison / Network integration / Protein interaction network

1 Introduction

One of major challenges for postgenomic biology is to understand how genes, proteins, and small molecules interact to form cellular systems [1–3]. It has been recognized that a complicated living organism cannot be fully understood by merely analyzing individual components, and that interactions of those components or networks are ultimately responsible for an organisms' form and functions. In recent years, with rapid progress of biological science, many high-throughput technologies have been developed for studying interactions of molecules, such as microarray, the two-hybrid assay, coimmunoprecipitation, and the chIP–chip approach, which can be used to screen for protein–protein interaction (PPI) [4] or to infer gene regulatory network (GRN) [5]. For instance, these technologies have been so far adopted to derive PPI networks for many model species [4], such as bacteria, yeast, nematode worm, and fruit fly. With increasingly accumulated data from those high-throughput technologies, molecular networks have been studied extensively from various aspects of living organisms. Those research works help biologists not only to understand complicated biochemical phenomena but also to elucidate the essential principles or fundamental mechanisms of cellular systems.

Correspondence: Professor Xiang-Sun Zhang, Chinese Academy of Sciences, Beijing 100080, China
E-mail: zxs@amt.ac.cn
Fax: +86-10-62561963

Abbreviations: DDI, domain–domain interaction; **GCEN,** gene coexpression network; **GO,** gene ontology; **GRN,** gene regulatory network; **MFGO,** modified and faster global optimization; **MN,** metabolic network; **PPI,** protein–protein interaction

* Additional corresponding author: Professor Luonan Chen
E-mail: chen@eic.osaka-sandai.ac.jp

From both biological and theoretical viewpoint, different interaction types among biomolecules form various types of networks or graphs, such as PPI networks, metabolic networks (MNs), signaling transduction networks, gene transcriptional networks, and GRNs. Detailed mathematical definition and biological meaning of these networks can be seen in a recent review [1]. In particular, the rapid development of complex network theory such as small-world property, right-skewed degree distribution, network transitivity property, network motif, and community structure also accelerate our understanding of topological structure of biological networks. Naturally, the well-understood graph-theoretical concepts can be used systematically to explore the topology, organization, function, and evolution of biological networks. Such studies will no doubt deepen our knowledge at a system-wide level and further enhance biological insights on living organisms.

In this paper, we aim to review the recent developments on topics related to molecular networks in a comprehensive manner. The contents of the survey cover global topological properties and local structural characteristics, network motifs, network comparison and query, detection of functional modules and network motifs, function prediction from network analysis, inferring molecular networks from experimental data as well as representative databases and software tools. As shown in Fig. 1 for the overview of the subjects in this paper, we mainly highlight those new advances of recent studies and also describe open challenging problems in this field.

2 Global measures of network topology

Topological analysis of biological networks is a topic of great interest in the field of current bioinformatics and systems biology, and provides quantitative insight into biological systems. A global measure of networks is introduced to characterize a network structure. There are four higher-level topological indices including average degree (K), clustering coefficient (C), average path length (L), and diameter (D), and four topological distributions including degree distribution $P(k)$, degree distribution of cluster coefficients $C(k)$, shortest path distribution SP(i), and topological coefficient distribution TC(k), which are of particular interest (see ref. [6–8] for details) and are comprehensively used in cellular networks, such as PPI networks [9, 10], MNs [11], gene coexpression

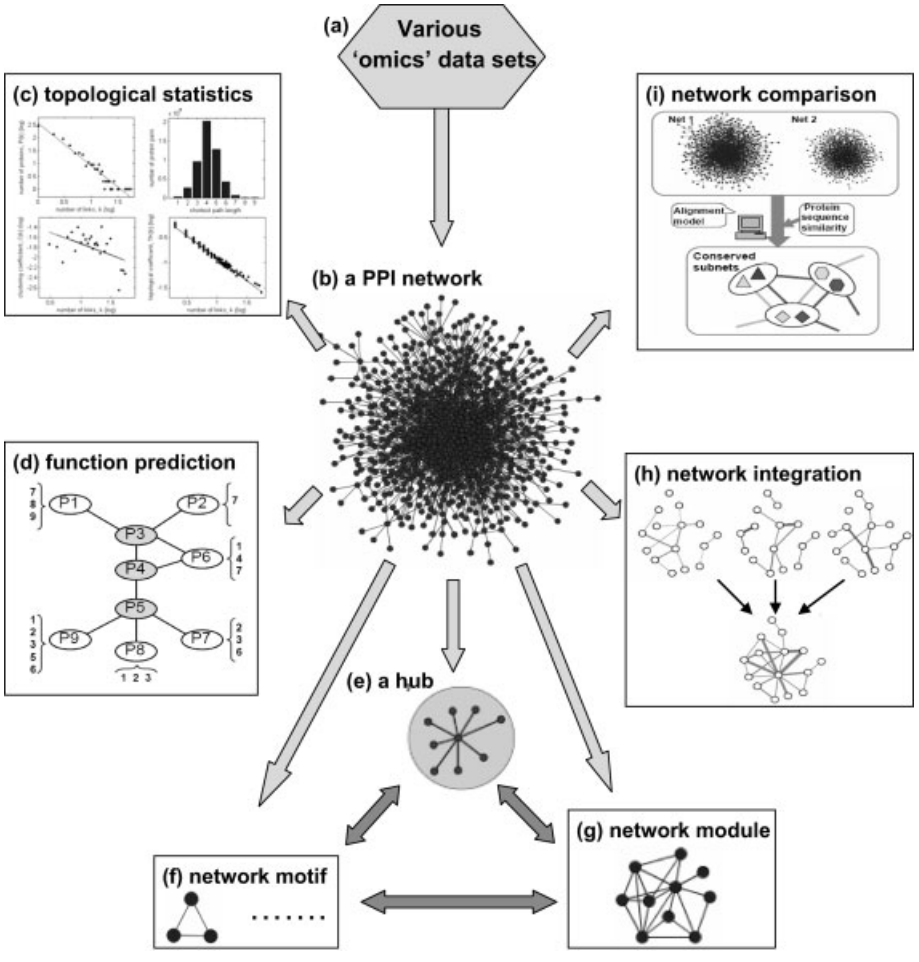

Figure 1. The overview of the subjects for biomolecular networks surveyed in this paper. (a) Many omics datasets provide materials to form various types of molecular networks or validate various studies of networks. (b) Molecular network–a PPI network. (c) Topologically statistical analysis, which shows four distributions of a PPI network. (d) Function prediction from network analysis, where only some of proteins have been annotated. (e) Hub proteins. (f) Network motifs. (g) Functional modules. (h) Network integration. (i) Network comparison and query.

networks (GCEN) [12], and domain interaction networks [13]. These criteria can efficiently capture the topological features of cellular networks and provide broadly insights into cellular evolution, molecular function, network stability, and dynamic responses [8, 11–13].

With ever increasing amount of available data on cellular networks, comparative analysis of cellular networks is expected to be an important step towards understanding the biological roles or functions of these networks and revealing evolutionary mechanisms at a system-wide level. In a recent study, Zhu and Qin [11] uncovered topological differences among the MNs of 11 single cell organisms by comparing those network structures. They showed that the design principles of Archaea MNs may be fundamentally different from those of bacteria and eukaryote and also that there are significant differences for functional mechanisms in three domains of life, by systematically comparing topological structures of MNs. Yook *et al.* [9] compared four available databases of the PPI networks for *S. cerevisiae* and found that each PPI network is of scale-free topology with hierarchical modularity. Recently, Li *et al.* [10] confirmed the similar conclusions in the PPI networks of *D. melanogaster* and *C. elegans*, and reported more topological features about these networks.

In general, biological networks have shown well scale-free or small-world property with hierarchical organization [8]. Such particular characteristics of network topology are also frequently been discovered in many nonbiological networks. These features provide possibility of exploring evolution or growth mechanisms of cellular networks, as indicated in ref. [1].

We should note that such observation is limited by the incompleteness and noise of biological data. Several types of molecular networks or cellular networks display scale-free topologies which are characterized by the power-law degree distribution. But Tanaka *et al.* argued that some PPI networks do not exhibit power-law statistics by evaluating degree sequences [14]. Another study suggested that the observed scale-free property of current PPI networks cannot be extrapolated to complete interactomes [15]. Moreover, such topological measures are only global and rough criteria, which are not able to distinguish more detailed topological structures. Taking Fig. 2 obtained from [16] as an example, there are two networks shown in (a) and (b) respectively, which are different. The network in (a) has a higher level of hierarchy. But the topological measures including degree distribution and clustering coefficient indicate that these two networks are topologically identical. In other worlds, many topological features may not be explicitly distinguished by these global measures.

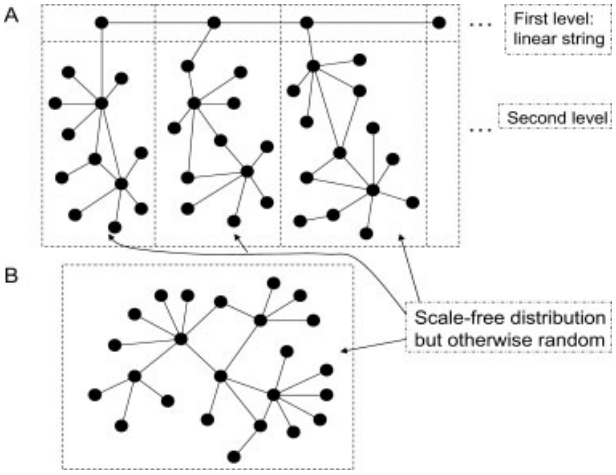

Figure 2. Two networks with identical measure of global topology. Network (a) has a higher level linear string of nodes which connect to the lower sets of subnets. These subnets follow the power law but are random in other respect just as network (b) has displayed.

3 Network centralities and hubs

To elucidate the fundamental mechanisms for the regulatory, interaction, metabolism, and transduction in a cell, a high-quality ranking or measure of network elements (*e.g.*, genes, proteins, transcripts, or metabolites) is essential. A network centrality is a quantitative measure of the position of an element (node) relative to other elements, and different centralities focus on different concepts or features. A network centrality can also be used to evaluate the relative importance of an element in the whole system. Actually central elements of molecular networks have been observed to be essential for many biological phenomena, such as viability, evolution, and stability. Fell and Wagner [17] showed that the most central metabolites are evolutionary conservative in MNs. In yeast PPI networks, Jeong *et al.* [18] uncovered that the centrality of a protein correlates with the essentiality of its gene by the observation that knocking out this gene would lead to a high probability of a lethal effect. Furthermore, by comparing six centrality measures, Estrada [19] showed that the subgraph centrality based on graph spectral properties of the network has the best performance in identifying essential proteins in the yeast PPI network. This study may have an important implication for selecting possible targets for drug discovery. Recently, a software tool called CentiBiN [20] for computation and exploration of centralities in molecular networks has been developed. The system computes 17 different centralities for directed or undirected networks, with which the importance of an element in molecular networks can be numerically evaluated.

Among those centralities, the degree centrality, possibly the simplest one, is defined as the degree of nodes, *i.e.*, the number of nodes that are linked with a node. In a molecular network, a node means a molecule, *e.g.*, a protein or a chemical. Since molecular networks may have well scale-free properties, biologists have special interest in nodes with high degrees, *i.e.*, so-called hubs [1]. Intuitively, one might expect to find some particularities of these proteins, which are actually observed to be essential [18] or to be slower evolving

[21]. Those hub proteins generally prefer to interact with lowly connected proteins, rather than to interact with other hub proteins in a PPI network [22]. For instance, the paper [12] showed that genes with high degrees tend to be essential and conserved in coexpression networks. Note that all these observations may be affected heavily by the noise data sources. Recently, Batada *et al.* [23] re-evaluated related problems by employing an extensive literature-curated dataset of PPIs, and argued that hubs are likely to be essential, or at least have a larger impact on fitness, but they are not slow evolving. He and Zhang [24] proposed an interesting alternative view on essential PPIs similar with essential proteins, *i.e.*, a small fraction of randomly distributed interactions are considered as essential, and each of those interactions is lethal to an organism when disrupted. Another interesting work is about "date" and "party" hubs investigated by Han *et al.* [25]. In PPI networks, party hubs are those hubs coexpressed with many of their neighbors, while date hubs are those not coexpressed with their neighbors. The authors suggested that these two types of hubs dynamically mediate modularity of networks. But a recent re-examined study, Batada *et al.* [26] argued some contrast observations that hub–hub interactions are not suppressed and a data hub does not have different biological attributes compared to other hubs, such as different rates of evolution. Ekman *et al.* [27] explored the characteristics between hubs and nonhubs, and date hubs and party hubs on domain level. They pointed that multiple and repeated domains are enriched in hub proteins, and long disordered regions important for flexible binding are statically significant in date hubs. Furthermore, Kim *et al.* [28] observed some evolutionary implications of 3-D structures with protein networks. However, although many biological phenomena are observed with the assumption of possible cellular mechanisms, we should note that all of those conclusions are not definite and depend on the quality and quantity of available data which are incomplete with biases.

4 Biologically significant subnets: Motifs and modules

Intuitively, a molecular network facilitates its functions through characteristic topological patterns. Various components of cellular networks including genes, proteins, and other molecules act in collaboration to carry out specific biological processes and biochemical activities, by forming relatively isolated functional units called modules in molecular networks. From the theoretical viewpoint, the decomposition of a large network into communities (or modules) is an effective way to gain insight or understand basic architecture of the complex network. On the other hand, a network motif is a significant recurring unit that may be a subunit of a module. Elucidating the essential roles of motifs and identifying modules in molecular networks are of great interests both theoretically and biologically. Figure 3 shows an example about motifs and modules.

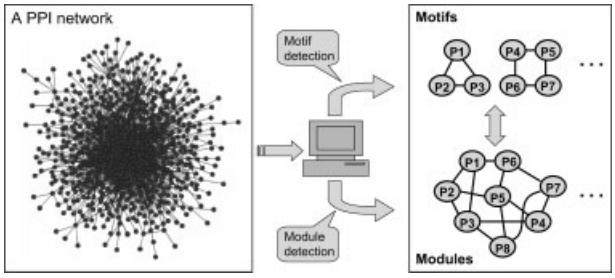

Figure 3. Motifs and modules in a PPI network. A representative module is extracted from a PPI network which corresponds well to a matched functional classification. Also two motifs are illustrated in the network and the module.

4.1 Network motifs

A network motif is a subunit (or subnet) of a complex network that is found to be significantly repeated, *i.e.*, these subnets appear significantly more frequent in the given network than expected by chance alone [29]. Such subnets are considered to be basic building blocks of many real complex networks including molecular networks. Obviously, enumerating all subnets with a given mode in a large network is practically infeasible. To overcome this difficulty, an algorithm based on a subnet importance sampling strategy has been developed. It can estimate densities of subnets, and detect motifs as a complexity of asymptotically independent of network size [30]. Moreover, efficient software tools for motif detection and visualization in large networks have also been developed [31, 32].

Motifs have been found and studied in various molecular networks, such as feed-forward loops and single-input motifs in transcriptional networks, and short cycles in PPI networks. Furthermore, many representative motifs were found in multiple organisms, which suggests the functional significance of network motifs for regulation, interaction, transduction, and evolution. For example, several identical motifs have been observed in the transcription regulatory network of diverse species. In particular, Zhu and Qin [11] found that *S. cerevisiae* and six bacterial species share identical three-node motifs and two four-node motifs in MNs, while four Archaeal species show great differences from them. Such a fact implies that Archaeal species may employ some different mechanisms to perform biological functions from bacterial species. By labeling the protein nodes in a PPI network with functional attributes of gene ontology (GO), Lee *et al.* [33] enumerated all recurring patterns of the annotated PPI network. They found that evolutionary constraints on the motifs have great differences from those with functional attributes. One can further explore structurally dynamical properties by checking distribution of motifs in different condition-specific cellular networks. It has been observed that the frequency of each motif type varies with different condition of the network. Luscombe *et al.* [34] built condition-specific transcriptional regulation network for yeast based on

transcriptional binding data and gene expression data. They observed that different frequencies of various regulatory motifs occurred depending on conditions of the network, which implies structurally dynamical changes in different conditions. By re-examining the datasets of Luscombe *et al.* [34], Zhang *et al.* [35] explored how differences in regulatory motif abundance are related to specific transcription hub factors which are considered to be crucial and essential in cellular networks [36]. Their studies show that different transcription hubs in a condition-specific network prefer different types of motifs, but variations in motif abundance cannot explain such preferences. They further pointed that motif preferences of transcription hubs change with variations of molecular networks.

These studies about motifs not only deepen the understanding of biological implications for network structures, but also help us gain insight into biological evolution as well as dynamical behaviors of living organisms. However, we should also note that motifs are not almighty for analyzing molecular networks. For instance, motifs cannot capture characteristic of connections among nodes. Moreover, Przulj *et al.* [37] argued that exploring the organization of infrequently appeared subnets is also important but enumerating all of subnets is computationally intense. Therefore, they designed efficient sampling heuristics to find small subnets in a PPI network by focusing only on specific parts of the network [38]. In contrast to network motifs emphasizing on statistical significance, one can estimate the distribution of different subnets and examine their global properties just like degree distribution [39], which can provide valuable information from other aspects of biological networks.

4.2 Functional modules in molecular networks

Biological networks have long been considered organized in a modular manner, which is composed of topologically or/and functionally relatively isolated subnetworks corresponding to specific biological units [40]. Generally, modules can be understood as subnets which are densely connected within themselves but sparsely connected with the rest of the network. Revealing modular structure in cellular networks is helpful for understanding biochemical processes and signal pathways. Note that module structure referred to as "community structure" in the field of complex system theory also has attracted great interests. A large number of computational methods for detecting community structures have been developed, and their detail descriptions can be found in a comprehensive review [41] and an evaluation paper [42].

As to molecular networks, several partition and local search methods have also been proposed to decompose the whole network into functional units. Specifically, hierarchical clustering methods have been proven to be an effective strategy for MNs and PPI networks. Ravasz *et al.* [43] analyzed the hierarchical structure of modularity in MNs, and authors of ref. [44–47] applied three different clustering methods respectively, based on different metrics induced by shortest-distance, graphical distances, and probabilistic functions, to investigate the module structure of the yeast PPI networks on a clustering tree. For example, Ravasz *et al.* [43] defined a simple (dis)similarity relationship between two nodes to cluster nodes in MNs. Arnau *et al.* [48] also applied the hierarchical clustering method based on the shortest path between any two nodes on PPI networks, where resampling strategy is adopted to overcome the low resolution problem for the shortest path measure due to the limited length. Lu *et al.* [47] suggested a simple measure to depict the link relationship between any two nodes. Diffusion kernel is also suggested to be employed as a universal similarity measure to construct the clustering tree [128]. On the other hand, although there are many (dis)similarity measures proposed so far for hierarchical clustering, it is lacking of a reliable way to evaluate those measures. For such a purpose, two evaluation schemes based on the depth of hierarchical tree and width of ordered adjacency matrix suggested by Lu *et al.* may be useful. Based on these two schemes, the preliminary results indicate that diffusion kernel outperforms other (dis)similarity criteria although a systematic evaluation should be conducted in a comprehensive manner.

Biological validation of the detected modules with biological data is essential. In general, functional homogeneity of proteins in a module with known function annotation from MIPS [49] or GO can be evaluated with a hypergeometrical distribution. Match between modules and experimentally determined complexes is always used to validate the biological significance of modules and to evaluate different detection algorithms [50–55]. The proteins included in the same module generally tend to share similar temporal expression profiles, subcellular localizations, and gene phenotypes, which support the functional relevance of modular organization.

There are still many remaining challenging problems to detect functional modules by computational methods in molecular networks. Specifically, the detected modules are often different from one method to another due to unclear boundaries between modules or hierarchical architecture of the network. Generally speaking, most methods introduce some parameters which are usually sensitive to local density of networks. Although we can learn the so-called optimal parameters relative to known function annotation or other genomic information, the results heavily depend on parameter tuning and the datasets. Evaluation of the algorithms for detecting functional module is also an intricate work. Recently, Brohée and Helden [56] made a comparative assessment of four algorithms including MCODE [50], SPC [52], RNSC [53], and MCL [57], for which they showed that MCL is remarkably robust to graph alterations and effective for the extraction of complexes. But it is only a small scale evaluation, a further assessment of existing methods including the recent ones [58–60] is required in a systematic and comprehensive manner.

There are two problems with respect to analysis of functional modules. A molecular network (*e.g.*, PPI network) generally is very sparse, but most methods only identify strongly connected subgraphs as modules. Hence, only a few modules can be detected, such as in ref. [50, 53], where only a few proteins have been covered [61]. Specifically, King *et al.* [53] detected merely about 200, 47, and 5 modules in the yeast, fly, and worm PPI networks respectively, by applying RNSC algorithm. Most current methods are partition algorithms which mean that each protein belongs to only one specific module. Such algorithms are not suitable for finding overlapping modules. Although some local search methods can detect modules with overlap, there is no detail discussion on the possible significance and biological implication of overlapping nodes. Recently, some methods [62–64] have been developed to detect fuzzy modules and the so-called CFinder method [62] has been applied to molecular networks [65]. But the CFinder method is too restrictive and its basic element is 3-clique, and thereby it can detect few modules in fly and worm PPI networks [61]. A relaxed overlapping functional module detection algorithm which cannot only detect abundant modules but also uncover meaningful overlapping nodes (proteins/genes or other molecules) is imperative so as to further improve the accuracy.

5 Function predicting from network analysis

A challenging problem in computational biology is to annotate proteins with biological functions. Various methods based on sequence similarity or structural similarity for proteins, phylogenetic profiles, gene expression data, protein complexes, and so on have been developed for this problem [66]. More recently, a new class of method has been proposed based on the network topology of the protein interaction networks from available interaction data, or from multiple data integration of protein interaction data and sequence similarity, protein structure data (or domain information), and gene expression data. The protein interaction data are directly related to network structures, thereby not only enlarging the information bases from the structural aspect but also enabling the application of new techniques of systems biology to function prediction. Now, systematic protein function classification databases with a specifically designed complex dynamic scheme have been constructed, such as MIPS and GO, which provide detailed annotations for growing number of proteins with a hierarchical organization. Naturally, how to exploit the global PPI network structure and comprehensive function information for predicting unknown proteins or unclassified proteins becomes an attracting problem. Figure 1(d) shows a toy PPI network, in which five proteins have known function annotations but other three have unknown functions. The earlier "Neighboring Counting" [67] and chi-square statistics [68] methods assign a protein with a function that is the most common or over-represented among its neighbors. More effective global optimization methods such as Markov Random Field [69, 70] and Monte Carlo simulated annealing (MCSA) [71] have been employed to predict protein functions by global topological link information including links among unclassified proteins with more promising results. Lanckriet *et al.* [72] introduced an integrated support vector machines classifier for function prediction, in which PPI data were used to derive one of the kernels using pairwise interaction similarity between proteins. Nabieva *et al.* [73] introduced a network-based algorithm that simulates functional flow between proteins. More recently, a modified and faster global optimization (MFGO) method [74] was designed to avoid the high intensive computation of the repeated process of MCSA. However, the iterative local optimization algorithm for MFGO can only give a local optimal solution due to the NP-hard nature of the problem. Furthermore, the authors also stated that MFGO can be used to predict multiple functions, however as a matter of facts, this is a hard problem because there is no knowledge or technique to determine an appropriate threshold in MFGO. Although these approaches demonstrated that the use of a variety of machine learning, statistical techniques, and global optimization can improve prediction performance, they bank on the same fundamental assumption that the interaction partners of a protein are likely to share similar functions with it. Since two proteins with indirect relationship or correlation, *i.e.*, the two proteins which interact with the same protein or proteins [75], also have a high probability to share same functions, Chua *et al.* [75] proposed a new function prediction method by employing the indirect functional association which has shown relative superiority over other methods. Although the topological link (direct or indirect) relationship between proteins implies their functional similarity, the results are biased (limited) by the noise and incomplete data. In a broad view, a method which can integrate PPI data with various genomic data to extend the coverage and reliability of data, is required so as to predict functions in a more accurate manner [66].

6 Network comparison

Molecular networks orchestrate the sophistic and complex functions of the living cells. Various organisms differ not only because of differences of constituting proteins, but also because of architectures of their molecular networks. Hence, it is essential to address the similarities and differences in the molecular networks by comparative network analysis, which can directly be applied for analyzing signal pathways, finding conserved regions, discovering new biological functions or understanding the evolution of protein interactions. The classical problem on this area includes network alignment, multiple network alignment and subnet query in a given network [76–80]. The problem of network alignment is to detect subnets that are conserved across species or within species by comparing two networks. Figure 1(i) shows a

representative example of network alignment, in which a detailed probabilistic model is used to identify conserved protein subnetworks across the species within the combined alignment network constructed based on protein interactions and sequence homologue information. Similarly, network querying is to search subnets which may be well-known functional units or specially interested in a given network (see Fig. 4 for illustration). Up to now, the researches have made great progress in this field. We list recent methods, each of which focuses on some specific network type or network comparison model in Table 1. As indicated in Table 1, various network alignment algorithms for molecular networks have been proposed in recent studies, which are either mainly based on sequence similarities, such as PathBLAST [81] and Local graph alignment algorithm [82], or mainly based on network architecture similarities, such as pairwise local alignment algorithm [83] and heuristic graph comparison algorithm [84]. But the PathBLAST [81] is heuristic and can only detect short (three or four nodes) linear paths, while the MetaPathwayHunter [78] method can be mainly applied to relatively small and simple tree-like networks. Except some special cases such as alignment of linear chains, the problem is computationally hard and similar with the generalized subgraph isomorphism under certain formulations. Due to the high complexity to compare such molecular networks, most of the conventional approaches either restrict comparative analysis to special structures, such as pathways without loops, or adopt heuristic (or approximate) algorithms due to computational burden. To overcome such difficulty for computational complexity, Berg and Läsig [85] proposed an evolution-based method for network alignment which estimates the relative weight of node and link similarity score systematically by a Bayesian parameter inference model. The method can be applied to both weighted and weighted undirected molecular networks. We also developed an alignment tool "MNAligner" [129] based on an integer quadratic programming model to align networks in an accurate and efficient manner. The method is rather general and can be applied not only to unweighted and undirected networks, but also to weighted and directed networks. NetworkBLAST [77] is a three-species network alignment platform which extends a likelihood-based scoring scheme, but scales poorly with the number of networks, *i.e.*, it is limited only to a few (five or four) networks. Moreover, an evolution-based scoring assessment function for existing approaches is lacking except MaWISH used in ref. [83]. Recently, a distinguished progress is that Flannick *et al.* [86] developed a robust and fast alignment tool "Grælin" for multiple large interaction networks. A pathway querying tool QPath [87] has also been developed for a linear query pathway with respect to a network of interest. The algorithm can efficiently search the network for homologous pathways, allowing both insertions and deletions of proteins in the identified pathways. Clearly, similar to the querying method of sequences, a universal querying system that can query an interesting small network with complex substructures

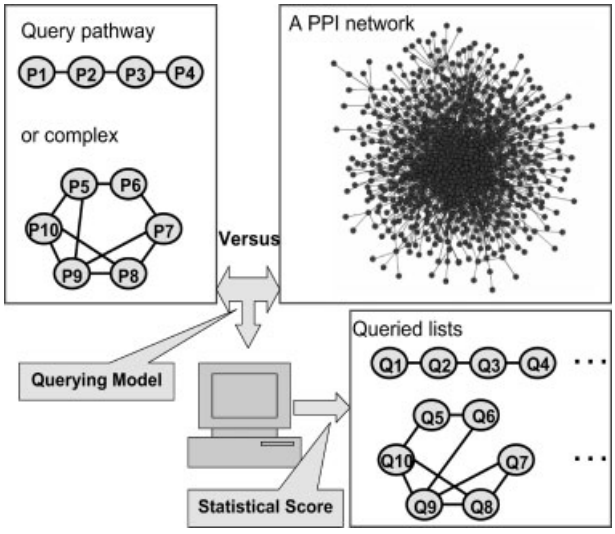

Figure 4. Network querying. The goal of network querying is to uncover subnets in a given network that are similar to the query subnet. Such subnet is previously known to be a functional unit, or one even known complex, or one particularly interest subnet.

(*e.g.*, protein complexes) efficiently in various types of molecular networks is highly demanded in the area of both bioinformatics and systems biology.

7 Inferring molecular networks

There are two major problems with experimental data generated by high-throughput experimental biotechnologies, *i.e.*, insufficient amount and low quality for the available data. So far various "omics" data have been adopted to infer molecular networks. Limited by the space, here we only survey recent works on inferring PPI networks at the domain level and inferring GRNs with a dynamical model, by stressing how to extend the coverage and reliability of data.

Existing methods for physical PPI prediction generally can be classified into two main groups [88], the structure-independent methods and the structure-based methods. Recently, another interesting class of algorithms based on statistic analysis is also adopted to discover protein interactions at the domain level. With training data, those methods first calculate the probability of each domain pair, and then predict PPI based on the domain–domain interaction (DDI) information. Figure 5 shows a schematic description of the underlying principle. Since DDI has a clear biological implication, it has been widely adopted to derive the PPI. To infer PPI networks at the domain level from experimental data, conventionally there are the association method [89], the likelihood method [90], linear programming based method [91] and the association numerical method [92]. An interesting evaluation study [93] on the correctness of those DDIs predicted by the association method [89], the likelihood

Table 1. Tools for network comparison. Ogata represents method developed by Ogata *et al.* [84]; Trusina represents method developed by Trusina [79]; Berg04 and Berg06 represent methods developed by Berg and Läsig in 2004 and 2006 respectively

Methods (net types)	Method description	Reference
Ogata (MN)	Single-linkage clustering is applied to a so-called network alignment graph which combines the genome ordering information and the network of successive enzymes in MN	[84]
PathBlast (PPI)	A dynamical programming algorithm is applied to search high-scoring paths in the alignment graph. By employing a likelihood-based scoring scheme, the method extends the original model to find conserved protein clusters	[4, 76, 81]
NetworkBlast (PPI)	A multiple network alignment framework which extends likelihood-based scoring scheme, and searches a network alignment graph of multiple networks	[77]
MaWISH (PPI)	This method is based on an evolution-based scoring scheme to detect conserved protein clusters	[83]
MetaPathwayHunter (MN)	The pathway alignment method is based on a subtree comparison algorithm. It has been developed into a network query procedure	[78]
Berg06 (GCEN)	The method is based on a scoring function measuring mutual similarities between two networks considering interactions patterns as well as node's sequence similarities. A Bayesian parameter inference is employed to obtain the optimal alignment	[85]
Trusina (GRN)	A simulated annealing algorithm is employed to minimize two local and global signaling scores which measure the difference of network architecture between two networks. The method is limited by the network size	[79]
Berg04 (GCEN)	A local graph alignment algorithm which is based on a scoring function measuring the statistical significance for families of mutually similar, but not necessarily identical subnets. The algorithm is conceptually similar to sequence alignment methodologies	[82]
QPath (PPI)	A network querying framework for identifying biologically significant pathways and inferring their function in protein interaction networks	[87]
Grælin (PPI)	Based on evolution-based probabilistic scoring scheme, the method is capable of searching large sets of dense networks for conserved subnets. The algorithm has adopted some strategy of sequence alignment	[86]
MNAligner (PPI, MN)	By employing both molecule similarity and architecture similarity, the method is based on an integer quadratic programming model to find the conserved substructures in an accurate manner	[129]

method [90], and their *p*-value method was carried out based on interacting protein pairs with crystal structure evidence. However, all those methods were outperformed by a random method for the largest group of protein pairs; such fact implies their potential shortcomings. There are also some studies [94–96] to integrate interactions from multiple organisms by extending the likelihood-based method. Recently, by exploiting special structures and composition of experimental data, we have developed a new association probabilistic method (APM) [97] based on domain interaction to infer PPIs, which outperforms other existing methods in terms of prediction quality and computational efficiency. APM can also be applied to multiple organisms' data by a simple manipulation. Przytycka's group [98] and our group [99] have independently introduced the concept of parsimony into explanation of PPI network with DDIs by a similar combinatorial optimization approach. This study hypothesizes that interactions between proteins are evolved and organized in a parsimonious way, which clearly explains the experimental data for protein interactions at the domain level [98, 99].

On the other hand, typically experimental data for PPIs suffer from high false positive and false negative rates [15], which significantly affect the accuracy and reliability of the prediction. Although integrating multiple genomic datasets

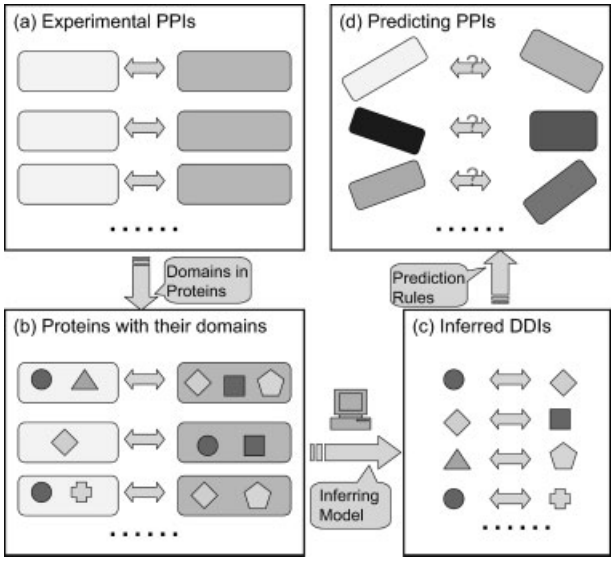

Figure 5. The schematic description of inferring PPIs at the domain level. (a) A list of experimentally determined PPIs. (b) The appropriate domains are assigned to each of the interacting proteins. (c) Inferred DDIs from PPIs and proteins' domain information. (d) Predicted PPIs from DDIs and their domain information.

from various organisms is one way to alleviate this problem [88], further research works are needed from both theoretical and experimental perspectives.

Large-scale microarray gene expression data provide new ways to learn gene regulation or construct GRNs or signal pathways [100]. As an important challenging problem of reverse engineering, a variety of models have been developed for this problem, such as the simple Boolean network model and dynamic Bayesian network model [100, 101], Gaussian graphical modeling [102], difference and differential equations model [103, 104], and so on. Specifically, the conventional Bayesian network model is based on directed acyclic graphs, including the stochastic nature of the considered biological processes, to identify the topology of GRNs. Recently, the Bayesian approach has been extended to the network with loops by adopting a dynamical architecture or a dynamical Beyesian network. A major problem for the Bayesian approaches is computationally intensive. On the other hand, the linear differential equations or liner dynamical models can also capture important features of a large-scale GRN [103]. However, as the same as most other problems of reverse engineering, even for a simple linear model, it suffers from curse of dimensionality problem, *i.e.*, the number of genes sampled in a microarray experiment is invariably much larger than the number of samples with the consequence that myriad networks can reproduce the observed data with fidelity. Efforts to constrain the model space by incorporating additional information from interventions and perturbations, other types of molecular data, or literature mining, work on small-scale systems but are still limited by increasing gene numbers. Alternative approaches make simplifying assumptions about network topology, or postulate that the microarray data are drawn randomly from a Gaussian distribution. Furthermore, similar to other problems in computational biology, all results heavily depend on the experimental data, which are not only extremely insufficient but also highly noisy. In contrast to the conventional methods that are mainly based on a single time-course dataset, we recently proposed a novel method GNRinfer [5] to infer GRN which exploits multiple time-course microarray datasets even with different conditions in a universal framework. The method theoretically ensures the derivation of the most consistent network structure with respect to all of the datasets, thereby not only significantly alleviating the problem of data scarcity but also remarkably improving the prediction reliability. But note that this model is based on an assumption that the structure of the regulatory network is stationary, and does not rewire under the environmental conditions for those different datasets. This means that the change of environmental conditions alters the level of gene expression instead of the network structure (actually most of methods use this assumption). Actually, GNRinfer can be further extended to identify the conserved network patterns or motifs from the datasets of either the same species or different species.

8 Network materials and software tools

A number of high-throughput biotechnologies have been applied to biological systems, which provide a huge amount of "interaction" maps for many organisms. Various omics data as well as their integration surveyed in an excellent review [3], provide great opportunity to uncover biological mechanisms. For instance, by applying a probabilistic model, Xia *et al.* [105] integrated 27 heterogeneous genomic datasets to predict PPI networks in human. Here, we collect the important resources for known PPI networks and MNs in Tables 2 and 3 respectively. Table 4 summarizes the representative software tools for network analysis.

Analyzing and modeling the molecular networks have become a central theme in systems biology. Therefore, versatile software tools for interactively displaying, analyzing, assimilating, modeling, integrating various biological networks are of great in demand. Currently, as shown in Table 5, a number of such tools have been developed by different groups, of which the most notable tools are Cytoscape [106], Osprey [107], Pajek [108], and so on. Cytoscape is an open source software project for integrating molecular interaction networks with various genomic datasets into a unified framework. The platform can be applied to any system of molecular components and interactions. Specially, it is most powerful when used in conjunction with large databases of various interaction types that are increasingly available for model organisms. The core of Cytoscape's software provides basic functionality to layout and query the network, to visually integrate the network with various genomic data

such as gene expression profiles and phenotypes, and to link the network to functional annotations' databases. Another distinguished characteristic is that the core is extensible through a straightforward plug-in architecture, which allows rapid development of additional computational analysis and features. For instances, GenePro [109], a new plug-in to the software, can greatly facilitate the visualization and analysis of PPI networks and the validation of various functional modules detection methods. The plug-in architecture has facilitated this field greatly and about 35 software tools as plug-in of Cytoscape (http://www.cytoscape.org/plugins2.php) have been developed until now. Also other tools such as graph drawing tools (*e.g.*, CADLIVE, CellDesigner,

Table 2. Resources for protein interaction data

WWW resources	Web sites
HPRD	http://www.hprd.org
BIND	http://www.bind.ca/
DIP	http://dip.doe-mbi.ucla.edu/
MINT	http://mint.bio.uniroma2.it/mint/
PIN	http://pin.mskcc.org/
STRING	http://string.embl.de/
HPID	http://wilab.inha.ac.kr/hpid/
IntAct	http://www.ebi.ac.uk/intact/index.jsp
OPHID	http://ophid.utoronto.ca/
Reactome	http://www.reactome.org/
MIPS/MPPI	http://mips.gsf.de/proj/ppi/
PPID	http://www.anc.ed.ac.uk/mscs/PPID/
BioCarta	http://www.biocarta.com
KEGG	http://www.genome.jp/kegg/
pSTIING	http://pstiing.licr.org/
GeneNet	http://wwwmgs.bionet.nsc.ru/mgs/gnw/genenet/
HiMAP	http://www.himap.org/

Table 3. Resources for metabolic pathways

WWW resources	Web sites
EMP	www.empproject.com/
DOCQS	http://doqcs.ncbs.res.in/doqcs/
Reactome	http://www.reactome.org/
BioCyc	http://www.biocyc.org/
MetaCyc	http://metacyc.org/
PANTHER	http://www.pantherdb.org/
KEGG	http://www.genome.jp/kegg/
Biozon	http://biozon.org/
BioCarta	http://www.biocarta.com/genes/index.asp
GenMAPP	http://www.genmapp.org/

Table 4. Software tools for network analysis

	Tools'name	Web sites
Global topological analysis		
	tYNA	http://networks.gersteinlab.org:8080/tyna/
	NetworkAnalyzer	http://med.bioinf.mpi-inf.mpg.de/netanalyzer/index.php
Network centralities		
	CentiBiN	http://centibin.ipk-gatersleben.de/
Network motif		
	Mfinder(MDraw)	http://www.weizmann.ac.il/mcb/UriAlon/groupNetworkMotifSW.html
	MAVisto	http://mavisto.ipk-gatersleben.de/
	FANMOD	http://www.minet.uni-jena.de/wernicke/motifs/
	Network Motif Finder	http://www.cytoscape.org/plugins2.php
Network modules (clustering)		
	MCODE	http://cbio.mskcc.org/bader/software/mcode/
	MCL	http://micans.org/mcl/
	SPC	http://www.vcclab.org/lab/spc/
	Edge-betweenness	http://jung.sourceforge.net/
	(PEBC)	http://www.cs.ucr.edu/qyang/software.html
	CFinder	http://www.cfinder.org/
	PRODISTIN	http://gin.univ-mrs.fr/webdistin
Network comparison (alignment/query)		
	PathBlast	http://www.pathblast.org/
	NetworkBlast	http://chianti.ucsd.edu/NetworkBlast/
	MaWISh	http://www.cs.purdue.edu/homes/koyuturk/mawish/
	MetaPathwayHunter	http://www.cs.technion.ac.il/olegro/metapathwayhunter/
	NetAlign	http://www1.ustc.edu.cn/lab/pcrystal/NetAlign/
	Grælin	http://graemlin.stanford.edu/

Table 5. Resources for network visualization and analysis

Softwares	Description	Web sites
Cytoscape	A bioinformatics software platform for visualizing molecular interaction networks and integrating interactions from other state data. A distinct feature is plugins which are available for network and molecular profiling analysis, with new layout, additional file format support, and connection to databases	http://www.cytoscape.org/
Osprey	A software platform for visualization and manipulation of complex interaction networks which builds data-rich graphical representations	http://biodata.mshri.on.ca/osprey/servlet/Index
Pajek	A universal platform for analyzing and visualization of large networks by implementing various graph approaches	http://vlado.fmf.uni-lj.si/pub/networks/pajek/default.htm
VANTED	A system for advanced data analysis and visualization in the context of biological networks	http://vanted.ipk-gatersleben.de
Gravisto	An editor for graphs and a toolkit for implementing graph visualization algorithms	http://gravisto.fmi.uni-passau.de/
Gaggle	An open-source software system for integrating bioinformatics software and data sources	http://gaggle.systemsbiology.net
GenePro	A cytoscape plug-in for advanced visualization and analysis of interaction networks	http://genepro.ccb.sickkids.ca
tYNA	A web tool for managing, comparing, and mining multiple networks	http://networks.gersteinlab.org/tyna
N-Browse	An interactive graphical browser for biological networks which aims to provide a simple, intuitive, and dynamic interface to molecular interaction data.	http://nematoda.bio.nyu.edu:8080/NBrowse/about.html
ProViz	A software platform for navigating in large graphs and exploring biologically relevant features. It adopts emerging standards such as GO and PSI-MI.	http://cbi.labri.fr/eng/proviz.htm
APID	An interactive bioinformatics web tool developed to integrate and analyze in a unified and comparative platform mainly for currently known information of PPIs.	http://bioinfow.dep.usal.es/apid/.
PPINA	A software that facilitates working with PPI networks by integrating data from multiple sources, providing a library that handles graph-related tasks	http://sbi.imim.es/piana
VisANT	A web-based software framework for visualizing and analyzing many types of networks of biological interactions and associations.	http://visant.bu.edu
Medusa	A Java application for visualizing and manipulating graphs of interaction	http://www.bork.embl.de/medusa
CNplot	A simple technique for the visualization of global connectivity within preclustered network data	http://csb.stanford.edu/nbatada/VCN.html
Biological-Networks	A software for visualizing molecular interaction networks, integrating with other network data such as ontologies and taxonomies, gene expression data, querying all types of data to extract meaningful relations, pathway modeling, and simulation	http://brak.sdsc.edu/pub/BiologicalNetworks
WebInter-Viewer	A software for visualizing and analyzing molecular interaction networks	http://interviewer.inha.ac.kr/
CADLIVE	A novel layout algorithm to draw complex biochemical networks	http://www.cadlive.jp/
CellDesigner	A process diagram editor for gene-regulatory and biochemical networks	http://www.systems-biology.org/
VitaPad	A visualization tools for the analysis of pathway data	http://bioinformatics.med.yale.edu
Cellware	A multialgorithmic software environment for modeling and simulating both deterministic and stochastic events in the cell	http://www.bii.a-star.edu.sg/sbg/cellware
NetBuilder	An interactive graphical tool for representing and simulating genetic regulatory networks in multicellular organisms.	http://strc.herts.ac.uk/bio/maria/NetBuilder/index.html
BioLayout	A platform of versatile network visualization of structural and functional relationships	http://www.biolayout.org/
E-Cell	An object-oriented software suite for modeling, simulation, and analysis of large scale complex systems such as biological cells	http://www.e-cell.org/software/e-cell-system
Jdesigner/SBW	An application which allows one to draw a biochemical network and export the network in the form of SBML	http://sbw.kgi.edu/software/jdesigner.htm
PATIKAweb	A web interface for retrieving and analyzing biological pathways and providing a user-friendly interface, dynamic visualization, and automated layout and so on	http://www.cs.bilkent.edu.tr/~patikaweb/
Graphviz	A way of representing structural information as diagrams of abstract graphs and networks	http://www.graphviz.org/

Jdesigner/SBW), cell dynamical simulation tools (*e.g.*, E-Cell and Cellware), multiple information integration tools (*e.g.*, ProViz, VisANT, BiologicalNetworks) have significantly enriched this field.

9 Conclusions

Instead of individual components, it is interactions of those components or networks that are ultimately responsible for an organisms' form and functions. The huge accumulation of various omics data greatly enriches the field of molecular networks. Naturally, network-based analysis of cellular networks can lead to deep insights into biological systems by discovering biological functions and revealing essential mechanisms at the molecular level in a system-wide manner. Here, we surveyed the related themes in this field with the special emphasis on the computational aspect. As a rapid ongoing field, further research works and efforts from both experimental and theoretical perspectives are expected to exploit the great potential on understanding fundamental mechanism of living organisms and challenging both biological and medical problems not at individual component level but at a system-wide level. Limited by the space, several related studies on molecular networks were not included into this survey. For instance, uncovering functional modules, network motifs, and new functions performed on multiple networks [110–115] are not included. In addition, other important topics related to dynamics of biological networks, such as designing and constructing biomolecular networks *in vivo* [116–119], modeling and Analyzing nonlinear molecular networks [120–122], cooperative behaviors of biological systems, and biomolecular communication *via* networks or signal pathways [123–127] are also not included in the review, and will be given in other publication in future.

This work is partially supported by the National Natural Science Foundation of China under grant no. 10631070, and the Ministry of Science and Technology, China, under grant no. 2006CB503905.

10 References

[1] Albert, R., *J. Cell. Sci.* 2005, *118*, 4947–4957.

[2] Lee, I., Date, S., Adai, A., Marcotte, E., *Science* 2004, *306*, 1555–1558.

[3] Joyce, A. R., Palsson, B. O., *Nat. Rev. Mol. Cell. Biol.* 2006, *7*, 198–210.

[4] Kelley, B. P., Sharan, R., Karp, R., Sittler, E. T. *et al.*, *Proc. Natl. Acad. Sci. USA* 2003, *100*, 11394–11399.

[5] Wang, Y., Joshi, T., Xu, D., Zhang, X.-S., Chen, L., *Bioinformatics* 2006, *22*, 2413–2420.

[6] Stelzl, U., Worm, U., Lalowski, M., Haenig, C. *et al.*, *Cell* 2005, *122*, 957–968.

[7] Yu, H., Zhu, X., Greenbaum, D., Karro, J., Gerstein, M., *Nucleic. Acids Res.* 2004, *32*, 328–337.

[8] Barabasi, A., Oltvai, Z., *Nat. Rev. Genet.* 2004, *5*, 101–113.

[9] Yook, S. H., Oltvai, Z. N., Barabasi, A. L., *Proteomics* 2004, *4*, 928–942.

[10] Li, D., Li, J., Ouyang, S., Wang, J. *et al.*, *Proteomics* 2006, *6*, 456–461.

[11] Zhu, D., Qin, Z. S., *BMC Bioinformatics* 2005, *6*, 8.

[12] Bergmann, S., Ihmels, J., Barkai, N., *PLoS Biol.* 2004, *2*, E9.

[13] Wuchty, S., Almaas, E., *BMC Evol. Biol.* 2005, *5*, 24.

[14] Tanaka, R., Yi, T.-M., Doyle, J., *FEBS Lett.* 2005, *579*, 5140–5144.

[15] Han, J.-D. J., Dupuy, D., Bertin, N., Cusick, M. E., Vidal, M., *Nat. Biotechnol.* 2005, *23*, 839–844.

[16] Valente, A. X., Cusick, M. E., *Nucleic Acids Res.* 2006, *34*, 2812–2819.

[17] Fell, D. A., Wagner, A., *Nat. Biotechnol.* 2000, *18*, 1121–1122.

[18] Jeong, H., Mason, S. P., Barabási, A. L., Oltvai, Z. N., *Nature* 2001, *411*, 41–42.

[19] Estrada, E., *Proteomics* 2006, *6*, 35–40.

[20] Junker, B. H., Koschützki, D., Schreiber, F., *BMC Bioinformatics* 2006, *7*, 219.

[21] Wuchty, S., *Genome Res.* 2004, *14*, 1310–1314.

[22] Maslov, S., Sneppen, K., *Science* 2002, *296*, 910–913.

[23] Batada, N. N., Hurst, L. D., Tyers, M., *PLoS Comput. Biol.* 2006, *2*, e88.

[24] He, X., Zhang, J., *PLoS Genet.* 2006, *2*, e88.

[25] Han, J.-D. J., Bertin, N., Hao, T., Goldberg, D. S. *et al.*, *Nature* 2004, *430*, 88–93.

[26] Batada, N. N., Hurst, L. D., Tyers, M., *PLoS Biol.* 2006, *4*, e317.

[27] Ekman, D., Light, S., Björklund, Å. K., Elofsson, A., *Genome Biol.* 2006, *7*, R45.

[28] Kim, P. M., Lu, L. J., Xia, Y., Gerstein, M. B., *Science* 2006, *314*, 1938–1941.

[29] Li, C., Chen, L., Aihara, K., *IEEE Signal Processing Magazine* 2007, *24*, 136–147.

[30] Kashtan, N., Itzkovitz, S., Milo, R., *Bioinformatics* 2004, *20*, 1746–1758.

[31] Schreiber, F., Schwobbermeyer, H., *Bioinformatics* 2005, *21*, 3572–3574.

[32] Wernicke, S., Rasche, F., *Bioinformatics* 2006, *22*, 1152–1153.

[33] Lee, W. P., Jeng, B. C., Pai, T. W., Tsai, C. P. *et al.*, *BMC Genomics* 2006, *7*, 89.

[34] Luscombe, N. M., Babu, M. M., Yu, H., Snyder, M. *et al.*, *Nature* 2004, *431*, 308–312.

[35] Zhang, Z., Liu, C., Skogerbo, G., Zhu, X. *et al.*, *PLoS Comput. Biol.* 2006, *2*, e47.

[36] Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X., Gerstein, M., *Trends Genet.* 2004, *20*, 227–231.

[37] Przulj, N., Corneil, D. G., Jurisica, I., *Bioinformatics* 2004, *20*, 3508–3515.

[38] Przulj, N., Corneil, D. G., Jurisica, I., *Bioinformatics* 2006, *22*, 974–980.

[39] Vazquez, A., Dobrin, R., Sergi, D., Eckmann, J.-P. *et al.*, *Proc. Natl. Acad. Sci. USA* 2004, *101*, 17940–17945.

[40] Hartwell, L. H., Hopfield, J. J., Leibler, S., Murray, A. W., *Nature* 1999, *402*, C47–C52.

[41] Newman, M. E. J., *Eur. Phys. J. B* 2004, *38*, 321–330.

[42] Danon, L., Duch, J., Diaz-Guilera, A., Arenas, A., *J. Stat. Mech.* 2005, P09008.

[43] Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., Barabási, A. L., *Science* 2002, *297*, 1551–1555.

[44] Brun, C., Chevenet, F., Martin, D., Wojcik, J. *et al.*, *Genome Biol.* 2003, *5*, R6.

[45] Rives, A. W., Galitski, T., *Proc. Natl. Acad. Sci. USA*, 2003, *100*, 1128–1133.

[46] Samanta, M. P., Liang, S., *Proc. Natl. Acad. Sci. USA*, 2003, *100*, 12579–12583.

[47] Lu, H., Zhu, X., Liu, H., Skogerbo, G. *et al.*, *Nucleic Acids Res.* 2004, *32*, 4804–4811.

[48] Arnau, V., Mars, S., Marín, I., *Bioinformatics* 2005, *21*, 364–378.

[49] Mewes, H. W., Frishman, D., Guldener, U., Mannhaupt, G. *et al.*, *Nucleic Acids Res.* 2002, *30*, 31–34.

[50] Bader, G. D., Hogue, C. W., *BMC Bioinformatics* 2003, *4*, 2.

[51] Bu, D., Zhao, Y., Cai, L., Xue, H. *et al.*, *Nucleic Acids Res.* 2003, *31*, 2443–2450.

[52] Spirin, V., Mirny, L. A., *Proc. Natl. Acad. Sci. USA*, 2003, *100*, 12123–12126.

[53] King, A. D., Pr·zulj, N., Jurisica, I., *Bioinformatics* 2004, *20*, 3013–3020.

[54] Dunn, R., Dudbridge, F., Sanderson, C. M., *BMC Bioinformatics* 2005, *6*, 39.

[55] Pereira-Leal, J. B., Enright, A. J., Ouzounis, C. A., *Proteins* 2004, *54*, 49–57.

[56] Brohée, S., van Helden, J., *BMC Bioinformatics* 2006, *7*, 488.

[57] Van Dongen, S., *Ph.D. Thesis*, University of Utrecht, The Netherlands 2002.

[58] Altaf-Ul-Amin, M., Shinbo, Y., Mihara, K., Kurokawa, K., Kanaya, S., *BMC Bioinformatics* 2006, *7*, 207.

[59] Chen, J., Yuan, B., *Bioinformatics* 2006, *22*, 2283–2290.

[60] Luo, F., Yang, Y., Chen, C. F., Chang, R. *et al.*, *Bioinformatics* 2007, *23*, 207–214.

[61] Zhang, S., Liu, H. W., Ning, X. M., Zhang, X. S., *Proceeding of Sixth IEEE International Conference on Data Mining-Workshops* 2006, 130–135.

[62] Palla, G., Derényi, I., Farkas, I., Vicsek, T., *Nature* 2005, *435*, 814–818.

[63] Reichardt, J., Bornholdt, S., *Phys. Rev. Lett.* 2004, *93*, 218701.

[64] Zhang, S., Wang, R. S., Zhang, X. S., *Physica A* 2007, *374*, 483–490.

[65] Zhang, S., Ning, X. M., Zhang, X. S., *Comput. Biol. Chem.* 2006, *30*, 445–451.

[66] Friedberg, I., *Brief. Bioinform.* 2006, *7*, 225–242.

[67] Schwikowski, B., Uetz, P., Fields, S., *Nat. Biotechnol.* 2000, *18*, 1257–1261.

[68] Hishigaki, H., Nakai, K., Ono, T., Tanigami, A., Takagi, T., *Yeast* 2001, *18*, 523–531.

[69] Letovsky, S., Kasif, S., *Bioinformatics* 2003, *19*, i197–i204.

[70] Deng, M., Zhang, K., Mehta, S., Chen, T., Sun, F., *J. Comput. Biol.* 2003, *10*, 947–960.

[71] Vazquez, A., Flammini, A., Maritan, A., Vespignani, A., *Nat. Biotechnol.* 2003, *21*, 697–700.

[72] Lanckriet, G. R., Deng, M., Cristianini, N., Jordan, M. I., Noble, W. S., *Pac. Symp. Biocomput.* 2004, 300–311.

[73] Nabieva, E., Jim, K., Agarwal, A., Chazelle, B., Singh, M., *Bioinformatics* 2005, *21*, i302–i310.

[74] Sun, S., Zhao, Y., Jiao, Y., Yin, Y. *et al.*, *FEBS Lett.* 2006, *580*, 1891–1896.

[75] Chua, H. N., Sung, W. K., Wong, L., *Bioinformatics* 2006, *22*, 1623–1630.

[76] Sharan, R., Ideker, T., Kelley, B., Shamir, R., Karp, R. M., *J. Comput. Biol.* 2005, *12*, 835–846.

[77] Sharan, R., Suthram, S., Kelley, R. M., Kuhn, T. *et al.*, *Proc. Natl. Acad. Sci. USA* 2005, *102*, 1974–1979.

[78] Pinter, R. Y., Rokhlenko, O., Yeger-Lotem, E., Ziv-Ukelson, M., *Bioinformatics* 2005, *21*, 3401–3408.

[79] Trusina, A., Sneppen, K., Dodd, I. B., Shearwin, K. E., Egan, J. B., *Plos Comput. Biol.* 2005, *1*, e74.

[80] Sharan, R., Ideker, T., *Nat. Biotechnol.* 2006, *24*, 427–433.

[81] Kelley, P. B., Yuan, B., Lewitter, F., Sharan, R. *et al.*, *Nucleic Acids Res.* 2004, *32*, 83–88.

[82] Berg, J., Läsig, M., *Proc. Natl. Acad. Sci. USA* 2004, *101*, 14689–14694.

[83] Koyutürk, M., Grama, A., Szpankowski, W., *RECOM, LNBI*, 2005 *3500*, 48–65.

[84] Ogata, H., Fujibuchi, W., Goto, S., Kanehisa, M., *Nucleic Acids Res.* 2000, *28*, 4021–4028.

[85] Berg, J., Läsig, M., *Proc. Natl. Acad. Sci. USA* 2006, *103*, 10967–10972.

[86] Flannick, J., Novak, A., Srinivasan, B. S., McAdams, H. H., Batzoglou, S., *Genome Res.* 2006, *16*, 1169–1181.

[87] Shlomi, T., Segal, D., Ruppin, E., Sharan, R., *BMC Bioinformatics* 2006, *7*, 199.

[88] András, S., Vera, G., Adrián, K. A., Jeffrey, S., *Phys. Biol.* 2005, *2*, S1–S16.

[89] Sprinzak, E., Margalit, H., *J. Mol. Biol.* 2001, *311*, 681–692.

[90] Deng, M., Mehta, S., Sun, F., Chen, T., *Genome Res.* 2002, *12*, 1540–1548.

[91] Hayashida, M., Ueda, N., Akutsu, T., *Bioinformatics* 2003, *19*, i58–i65.

[92] Hayashida, M., Ueda, N., Akutsu, T., *Genome Inform.* 2004, *15*, 56–68.

[93] Nye, T. M., Berzuini, C., Gilks, W. R., Babu, M. M., Teichmann, S. A., *Bioinformatics* 2005, *21*, 993–1001.

[94] Liu, Y., Liu, N., Zhao, H., *Bioinformatics* 2005, *21*, 3279–3285.

[95] Riley, R., Lee, C., Sabatti, C., Eisenberg, D., *Genome Biol.* 2005, *6*, R89.

[96] Lee, H., Deng, M., Sun, F., Chen, T., *BMC Bioinformatics* 2006, *7*, 269.

[97] Chen, L., Wu, L.-Y., Wang, Y., Zhang, X.-S., *Proteins* 2006, *62*, 833–837.

[98] Guimarães, K. S., Jothi, R., Zotenko, E., Przytycka, T. M., *Genome Biol.* 2006, *7*, R104.

[99] Zhang, X.-S., Wang, R., Wu, L.-Y., Zhang, S., Chen, L., *IOMP Proceedings* 2006, *2*, 181–184, Springer-Verlag, Berlin 2006.

[100] Friedman, N., *Science* 2004, *30*, 799–805.

[101] Beal, M. J., Falciani, F., Ghahramani, Z., Rangel, C., Wild, D. L., *Bioinformatics* 2005, *21*, 349–356.

[102] Schafer, J., Strimmer, K., *Bioinformatics* 2005, *21*, 754–764.

[103] Gustafsson, M., Gustafsson, M., Hornquist, M., Lombardi, A., *IEEE/ACM Trans. Comput. Biol. Bioinform.* 2005, *2*, 254–261.

[104] Chen, L., Aihara, K., *IEEE Trans. on Circuits Syst. I* 2002, *49*, 602–608.

[105] Xia, K., Dong, D., Han J.-D. J., *BMC Bioinformatics* 2006, *7*, 508.

[106] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S. *et al.*, *Genome Res.* 2003, *13*, 2498–2504.

[107] Breitkreutz, B. J., Stark, C., Tyers, M., *Genome Biol.* 2003, *4*, R22.

[108] Batagelj, V., Mrvar, A., in: Jünger, M., Mutzel, P. (Eds.), *Graph Drawing Software*, Springer 2003, pp. 77–103.

[109] Vlasblom, J., Wu, S., Pu, S., Superina, M. *et. al.*, *Bioinformatics* 2006, *22*, 2178–2179.

[110] Koyuturk, M., Grama, A., Szpankowski, W., *Bioinformatics* 2004, *20*, i200–i207.

[111] Hu, H., Yan, X., Huang, Y., Han, J., Zhou, X. J., *Bioinformatics* 2005, *21*, i213–i221.

[112] Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S. *et al.*, *Proc. Natl. Acad. Sci. USA* 2004, *101*, 5934–5939.

[113] Moon, H. S., Bhak, J., Lee, K. H., Lee, D., *Bioinformatics* 2005, *21*, 1479–1486.

[114] Chen, Y., Xu, D., *Nucl. Acids Res.* 2004, *32*, 6414–6424.

[115] Zhang, L. V., King, O. D., Wong, S. L., Goldberg, D. S. *et al.*, *J. Biol.* 2005, *4*, 6.

[116] Hasty, J., Dolnik, M., Rottschäer, V., Collins, J. J., *Phys. Rev. Lett.* 2002, *88*, 148101.

[117] Wang, R., Zhou, T., Jing, Z., Chen, L., *Syst. Biol.* 2004, *1*, 71–84.

[118] Kobayashi, T., Chen, L., Aihara, K., *J. Theor. Biol.* 2002, *221*, 379–399.

[119] Chen, L., Wang, R., *IEEE Trans. on Circuits Syst. I* 2006, *53*, 2444–2450.

[120] Goldbeter, A., *Proc. R. Soc. Lond. B* 1995, *261*, 319–324.

[121] Chen, L., Wang, R., Kobayashi, T., Aihara, K., *Phys. Rev. E* 2004, *70*, 011909.

[122] Li, C., Chen, L., Aihara, K., Logic. *IEEE Trans. on Circuits Syst. I* 2006, *53*, 2451–2458.

[123] McMillen, D., Kopell, N., Hasty, J., Collins, J. J., *Proc. Natl. Acad. Sci. USA* 2002, *99*, 679–684.

[124] Li, C., Chen, L., Aihara, K., *BMC Syst. Biol.* 2007, *1*, 6.

[125] Li, C, Chen, L., Aihara, K., *PLoS Comput. Biol.* 2006, *2*, e103.

[126] Chen, L., Wang, R., Zhou, T., Aihara, K., *Bioinformatics* 2005, *21*, 2722–2729.

[127] Zhou, T., Chen, L., Aihara, K., *Phys. Rev. Lett.* 2005, *95*, 178103.

[128] Zhang, S., Ning, X. M., Zhang, X. S., *Eur. Phys. J. B* 2007, *57*, 67–74.

[129] Li, Z., Zhang, S., Wang, Y., Zhang, X. S., Chen, L., *Bioinformatics* 2007, *95*, doi: 10.1093/bioinformatics/btm156.

