# Systems-level insights into cellular regulation: inferring, analysing, and modelling intracellular networks

C. Christensen, J. Thakar and R. Albert

**Abstract:** Genes and gene products interact on several levels, forming transcriptional regulatory-, protein interaction-, metabolic- and signal transduction networks. Genetic, biochemical and molecular biology techniques have been used for decades to identify biological interactions; newly developed high-throughput methods now allow for the construction of genome-level interaction maps. In parallel, high-throughput expression data paired with computational algorithms can be used to infer networks of interactions and causal relationships capable of producing the observed experimental data. Graph-theoretical measures and network models are more and more frequently used to discern functional and evolutionary constraints in the organisation of biological networks. Perhaps most importantly, the combination of interaction and expression information allows the formulation of quantitative and predictive dynamic models. Some of the dominant experimental and computational methods used for the reconstruction or inference of cellular networks are reviewed, also the biological insights that have been obtained from graph-theoretical analysis of these networks, and the extension of static networks into various dynamic models capable of providing a new layer of insight into the functioning of cellular systems is discussed.

## 1 Definition of cellular networks

A system of elements that interact or regulate each other can be represented by a mathematical object called a graph (or network) [1]. At the simplest level, the system's elements are reduced to graph nodes (also called vertices) and interactions are reduced to edges connecting pairs of nodes. Edges can be either directed, specifying a source (starting point) and a target (endpoint), or non-directed. Thus, the interactions among genes and gene products can be depicted graphically as networks with either directed or non-directed edges. Directed edges are suitable for representing the flow of material from a substrate to a product in a reaction or the flow of information from a transcription factor to the gene whose transcription it regulates. Non-directed edges are used to represent mutual interactions such as protein–protein binding. Graphs can be augmented by assigning various attributes to the nodes and edges: multi-partite graphs allow representation of different classes of node (such as mRNA and protein), and edges can be characterised by signs (positive for activation, negative for inhibition), or weights quantifying confidence levels, strengths or reaction speeds.

In this review, we will focus on graphs constructed from molecular-to-cellular level interactions. At the genomic level, transcription factors activate or inhibit the transcription of genes to give mRNA. Since these transcription factors are themselves products of genes, the ultimate effect is that genes regulate each other's expression as part of gene regulatory networks. Similarly, proteins participate in diverse post-translational interactions that lead to modified protein functions or they bind to form protein complexes: the totality of these processes is called a protein–protein interaction network. The biochemical reactions of cellular metabolism can likewise be integrated into a metabolic network whose fluxes are regulated by enzymes catalysing the metabolic reactions. In many instances these different levels of interactions are integrated, as is done, for example, when the presence of an external signal triggers a cascade of interactions that involves both biochemical reactions and transcriptional regulation.

Genome-level information concerning cellular networks is often described using five 'omes': genome (DNA sequence information), transcriptome (the totality of transcribed genes), proteome (all the proteins in a cell), metabolome (all the metabolites in a cell), and interactome (the totality of protein interactions). Though all the 'omes', with the exception of the interactome, essentially provide vertex information, metabolome and transcriptome data can be used to infer networks of indirect interactions. During the last decade, the respective omics have produced an incredible quantity of molecular expression and interaction data, contributing to maps of specific cellular networks [2, 3].

The full representation of transcriptional regulatory maps associates two separate node classes with transcription factors and mRNAs, respectively, and has two types of directed edge, which correspond to transcriptional regulation and translation [4]. The edges describing transcriptional regulation can have two regulatory effects (signs): activation and inhibition. In protein interaction graphs, the nodes are proteins, and two proteins are connected by a non-directed edge if there is strong evidence of their association.



doi:10.1049/iet-syb:20060071

Paper first received 9th October and in revised form 30th December 2006

The authors are with the Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA

R. Albert is also with the Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802, USA

E-mail: cpc146@phys.psu.edu

A metabolic network can be represented as a directed and weighted tri-partite graph, whose three types of node are metabolites, reactions, and enzymes, and whose two types of edge represent mass flow and catalytic regulation, respectively (see Fig. 1). Fig. 1*a* shows the most detailed picture includes three types of node: reactants (circles), reactions (ovals) and enzymes (squares) and two types of edge, corresponding to mass flow (solid lines) or catalysis (dashed lines). The edges are marked by the stochiometric coefficients of the reactants. The metabolite network (Fig. 1*b*) shows that all the reactants that participate in the same reaction are connected by non-directed edges, and thus the network is composed of a set of completely connected subgraphs (triangles in this case). The reaction network (Fig. 1*c*) shows two reactions that are connected by a non-directed edge if they share a reactant. A similar graph can be constructed for the enzymes.

Mass flow edges connect reactants to reactions and reactions to products, and are marked by the stoichiometric coefficients of the metabolites [5, 6]; enzymes catalysing the reactions are connected by regulatory edges to the nodes signifying the reaction [7]. Several simplified representations include the substrate graph, whose nodes are reactants, joined by an edge if they occur in the same chemical reaction [8], and the reaction graph, whose nodes are reactions, connected if they share at least one metabolite.

Signal transduction networks, connecting extracellular signal inputs to the control of transcription factors, share a significant overlap with protein interaction networks and metabolic networks, as they involve both protein interactions and biochemical reactions. The nodes of signal transduction networks can be categorised by the function of the corresponding protein or molecule, and the edges are mostly directed, indicating the direction of signal propagation. Finally, information on gene co-expression [9], gene co-occurance [10], or genetic interactions [11] can be used to construct networks of gene functional relationships. For example, a synthetic lethal interaction, connecting a pair of genes whose combined knock-out causes cell death, indicates that these genes buffer for one another (see Fig. 2). The hypothetical cellular network module in Fig. 2 receives exogenous signals through the central top node, and the transfer of information (denoted by an arrow) to its sink node (bottom centre node) determines the response to the signal (or the phenotype). The nodes B–F form two synergistic functional complexes. The large (source and sink) nodes of this network correspond to essential genes. All other nodes represent non-essential genes, as there are two node-independent (redundant) pathways between the source node and sink node that can compensate for each other. Synthetic lethal interactions (paler lines) are indicated by pairs of nodes whose loss causes the disconnection of the source and sink nodes. Genetic interactions reflect a complex functional compensatory relationship and not a physical interaction [11].

This review focuses on four major questions. First, how can one map or infer the regulatory network underlying a biological system, and at what scale should such a map be constructed for an optimal understanding of the system? Second, what are the best measures that capture the most salient features of a biological network? Third, what are the selective constraints that determine how biological networks evolve? And finally, how does the topology of biological systems influence their dynamics and function? We aim to present a representative (while by necessity not comprehensive) picture of recent progress made in answering these questions. Several excellent books [12, 13] and reviews [14–16] offer a complementary reading on the state of the art in systems biology.

## 2 Experimental methods to detect interactions

Advances in molecular biology techniques increased the resolution to which interactions could be detected and gave flexibility to the implementation of other experimental techniques. More recently, in tandem with improvements in computational techniques for the analysis of complex systems, improved experimental methods capable of detecting a large number of interactions among biomolecules have produced high-throughput data. These improvements promise to change the focus of cell biology from an understanding of local, binary interactions to an understanding of the aggregation of these interactions into a functional system. In this section we review some of the methods currently used to obtain binary interaction data, and we examine the potential of these methods to produce high-throughput data.

### *2.1 Transcriptome data*

Transcriptome data convey the identity of each expressed gene and its level of expression (the abundance of its transcribed mRNA) for a defined population of cells. High-throughput mRNA data can be obtained by different approaches: for example, serial analysis of gene expression (SAGE) [17], hybridisation to small, high-density arrays containing synthetic oligonucleotides [18], and the most frequently used method, DNA microarrays [19, 20]. While the transcriptome

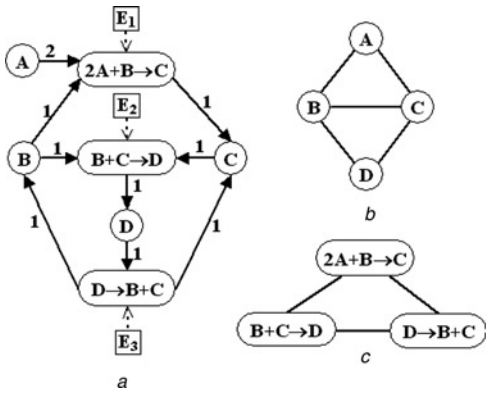



**Fig. 1** *Three possible representations of a reaction network with three enzyme-catalysed reactions and four reactants*

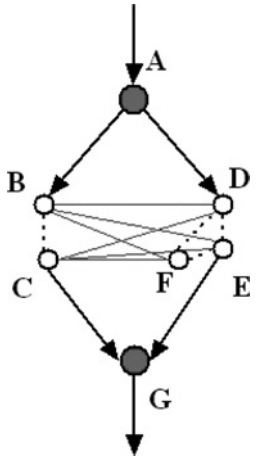



**Fig. 2** *Connections between pathway redundancy and synthetic lethal interactions*

constitutes node (component)-level information without information on interactions, it can be used to generate networks of functional relationships between genes. Gene expression data studied across time can be combined with computational methods to extract interactions [21, 22] (see Section 3).

The construction of transcriptional regulatory networks requires the large-scale detection of transcription factor (TF)–DNA interactions. Detection of a TF-bound nucleotide sequence confirms a TF–DNA binary interaction. Once the nucleotide sequence is known, array technology is used to look for the binding of other known TFs to the same site. The protocol for the detection of nucleotide sequences entails isolating the specific nucleotide sequence, followed by its analysis and identification.

Methods for TF binding-site isolation include DNA footprinting and chromatin immunoprecipitation (ChIP), with which the bound TF protects a target DNA segment from DNase-instigated degradation or shearing [23, 24]. For ChIP the protein and the bound DNA fragments are precipitated using a specific antibody against the TF [25]. The ChIP method has been used in recent high-throughput studies investigating 106 TFs [4] and 11 TFs that function at the G1/S transition [26]. A third method is based on the fact that DNA around the binding sites is preferentially methylated by DNA adenine methyltransferase [27]. Subsequently, the non-methylated DNA fragments are digested and removed by specific enzymes [28, 29]. Sun *et al.* [29] used this technique to map protein–DNA interactions at high resolution along large segments of *Drosophila melanogaster* genomic DNA.

After the binding sites have been isolated, they are analysed with electrophoresis methods or autoradiography. In most cases, the concentration of the sequence is below detection level, and hence polymerase chain reaction (PCR) is used for amplification. The ChIPchip technique combines the ChIP technique with DNA microarray technology [26, 30]: thousands of DNA fragments are first purified via the ChIP method, and are then identified simultaneously with microarray experiments. Known information about binding site sequences can be retrieved by web-based tools that query databases such as the transcription factor database (TRANSFAC) [31], regulon database (RegulonDB) [32], and the Kyoto encyclopedia of genes and genomes (KEGG) [33].

### 2.2 Interactome data

The interactome represents protein–protein interactions, including the formation of enzyme complexes (which, in turn, direct biochemical interactions, post-translational modifications etc.) and interactions involved in signal propagation.

Protein precipitation methods (pull-down assays [34]) (Fig. 3) are most commonly used for detecting these interactions. Potential interactors can be targeted with native, processed or post-translationally modified proteins (baits) [35]. For high-throughput experiments, the bait is tagged by Hemaglutinin (HA), TEV (tobacco etch virus) [36] or a tandem affinity purification (TAP) tag (which consists of two IgG binding domains of *Staphylococcus aureus* protein A and a calmodulin binding peptide separated by a TEV protease cleavage site) [37]. For metazoans, tagging is combined with siRNA-mediated downregulation of the endogenous form of the bait protein [38]. The bait is allowed to interact with potential interactors (in for instance cell lysates), and it is then recovered with antibodies against the tag bound to a solid support along with its interacting partner. The final step in protein purification often involves

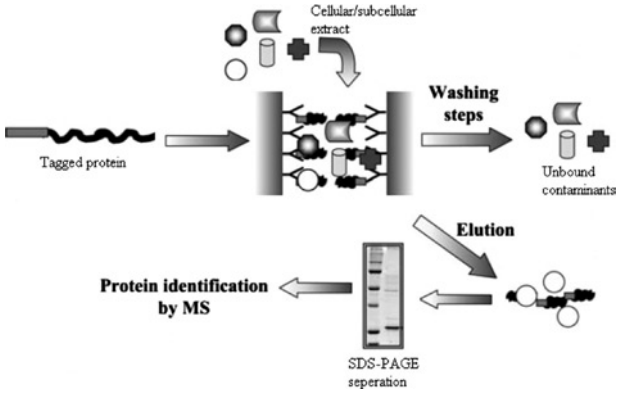



**Fig. 3** *Steps in a protocol of a traditional protein pull-down assay. Reprinted from [34] with permission from Springer Science and Business Media*

using protease cleavage (e.g. by Trypsin) to release the complexes from the support. The recovered sample is passed over a column (e.g. calmodulin), and is then eluted (e.g. with $Ca^{2+}$ chelators). This two-stage purification ensures low background noise and correspondingly high sample purity, but risks losing weak interacting partners or complex components due to the harsh purification procedure. The recently introduced protein chip method combines pull-down assay technology with microarray techniques and is used for large-scale studies [39].

The above methods capture only those interactions that have a certain minimum affinity and that occur in undisturbed cells. An often-used alternative, the yeast two hybrid (Y2H) method, exploits the interaction between the DNA binding (DB) domain and the activation domain (AD) of the yeast transcription factor GAL4p to detect binary interactions. Hybrid proteins are created by fusing the two proteins or domains of interest (generally called 'bait' and 'prey') to the DB and AD regions of Gal4p. These two hybrid proteins are introduced into yeast, and if transcription of Gal4p-regulated reporter genes is observed, the two proteins of interest are deemed to have formed an interaction, thereby bringing the DB and AD domains of Gal4p together and reconstituting the functional transcriptional activator.

Various mass spectroscopic (MS) methods, such as matrix assisted laser desorption/ionisation (MALDI), electrospray ionisation (ESI), or mass fingerprinting [40], are used to sequence proteins in a complex. The high accuracy of MS spectra, combined with knowledge of the genomic sequence of the organism in question, permits rapid and accurate identification of the proteins involved in the recovered complex. Quantitative MS can be used to detect low affinity (specific) interactions. The bait and a closely related, but binding-deficient, control bait are exposed to normal and isotopically labeled cell lysates, respectively. The associated proteins are mixed and analysed; the quantitative ratio of the label recovered by the bait and its close relative indicates the degree of exclusive binding to the bait. Non-specifically binding proteins are identified by their failure to show a significant quantitative difference in the ratio [41]. This method has been applied to identify signal-dependent, phosphorylation-mediated interactions [41]. In another adoption of the same principle, synthetic peptides were used as baits to selectively capture interactions of phosphopeptides, while the non-phosphorylated peptide acted as the control [42].

In two independent large-scale studies, hundreds of yeast proteins were tagged either by a small epitope tag [43] or by

the TAP tag [44]. Besides yielding interactions for about 25% of the yeast proteome, this study also revealed a higher-order organisation of complexes, defined via their shared components in different pull-downs. Y2H yielded the first interaction maps of metazoans, namely *D. melanogaster* [45], *Caenorhabditis elegans* [46] and *Homo sapiens* [47]. The density of biologically relevant interactions (true positives) in the highest-confidence subset of these maps was estimated at around 40–90%. Protein chips were used to assess the yeast proteome [48] by using tagged calmodulin bound to several wells, recapitulating known interaction partners and revealing many novel ones. The protein chip technique is especially suited to determining peptide or protein-domain interactions [49] and, in combination with Y2H data, has been used to establish interaction partners of the entire repertoire of yeast.

Once proteins have been identified, proteome and interactome databases can be mined for available information on interaction partners. Many interactome databases are maintained, including the database of interacting protein (DIP) [50], the biomolecular interaction network (BIND) [51], that of the Munich Information Center for Protein Sequences (MIPS) [52], the human protein reference database (HPRD) [53], and the yeast proteome database (YPD) [54].

### 2.3 Metabolome data

A substantial portion of the genome encodes enzymes that interconvert metabolites, synthesise cofactors, or regulate small molecule metabolism. Metabolites can, in turn, control gene expression and are allosteric regulators of enzymes. Experimental techniques of metabolome analysis include methods to trace and identify metabolites and to characterise enzymes catalysing reactions.

To characterise enzymes, either gene modification or enzyme isolation/purification methods are used. For example, Bussow *et al.* [55] prepared a set of *Escherichia coli* strains expressing tagged proteins in a 96-well format, spotted them onto a filter, and screened them for glyceraldehyde-3-phosphate dehydrogenase activity [56]. High-throughput methods of gene modification include the use of mutagens, followed by PCR amplification and identification [57]. Protein and DNA microarrays are also used to identify enzyme-encoding genes. The intracellular concentration of metabolites can reveal the activity of metabolic enzymes. Quantification of the change of several metabolite concentrations relative to the concentration change of one selected metabolite can reveal the site(s) of action in the metabolic network [58]. In the same way, comprehensive analyses of metabolite concentrations in mutants, providing 'metabolic snapshots', can reveal functions when snapshots from strains deleted for unstudied genes are compared to those deleted for known genes.

Metabolite identification necessitates the location and extraction of metabolites. Isotopic labelling (C-14, C-13) is used to trace carbon flow in a pathway; in *E. coli* 100 metabolites were identified using this technique [59]. Isotopic labelling can be used in conjunction with measurement of natural isotope levels to estimate fluxes [60]. Classical liquid–liquid extraction of cellular material is often used to extract phospholipids. Depending on the classes of metabolites (their mass range, thermostability and volatility) to be separated, different types of chromatography and MS techniques can be used. Field desorption was the first MS method used to analyse intact phospholipids; MALDI and electrospray ionisation (ESI), are used for high-throughput studies. The major advantages of ESI-MS are high accuracy, sensitivity, reproducibility, and the applicability of the technique to complex phospholipid solutions. Recently, Soga and colleagues developed a powerful analytical method using capillary electrophoresis–electrospray ionisation mass spectrometry that dramatically increases the number of metabolites that can be measured simultaneously [61, 62].

For flux quantification, MS [63, 64] combined with the separation ability of gas chromatography has been used for many years to measure the mass isotopomer distribution of intracellular metabolites in cell lysates for flux quantification in the context of disease diagnosis [65, 66]. Various other networks were analysed in subsequent studies using the same method [67, 68]. Other parameters that can be estimated in metabolic pathways are half-lives and turnover rates. The *in vivo* method includes intravenous injection of a labelled substrate that is integrated in a well defined metabolic pathway to determine its plasma input function to the tissue by following the tracer over time using positron emission tomography (PET) or autoradiography [69]. The function and metabolic pathways of identified metabolites can be looked up in the Encyclopedia of Metabolic Pathways (MetaCyc) [70] and the KEGG [33].

## 3 Computational methods for inferring network structure

Computational inference (also referred to as reverse engineering) aims to extract causal relationships from transcriptome, proteome and metabolome data. Inference of cellular networks allows for a clearer comprehension of the inner machinery of the cell, and when combined with modelling, can also be used to make experimentally verifiable predictions about cellular networks. A variety of computational methods for network inference exists; choosing a specific computational method depends on the nature of the data from which inferences will be made, on the type of network under consideration, on the features of the system one would most like to illuminate, and on the amount of computational time available to the researcher. Loosely speaking, computational methods fall into one of three classes: probabilistic, deterministic, and probabilistic/deterministic hybrids. In this section, we examine the dominant methods available in each of these classes, and we describe the circumstances under which a choice of method from a particular class is appropriate.

### 3.1 Probabilistic methods

Probabilistic methods, including clustering analyses [71, 72], data-mining [73, 74], and naïve Bayesian networks [75, 76], are applicable both to the inference of protein–protein interactions, and to the inference of functional relationships among genes based on similarities in gene expression profiles. In clustering algorithms, the correlation of the expression profiles of two target genes or proteins is scored against their individual correlations with all other profiles in the data set [72]. Genes or proteins with statistically similar profiles are then clustered using, for example, hierarchical clustering algorithms [21], self-organising maps [77], or K-mean clustering algorithms [78]. Recently-developed methods use topological measures to establish correlation thresholds for determining whether or not pairs of genes are co-expressed: transitivity criteria lead to excellent agreement with *Bacillus subtilis* operon structure and differential regulation [79], and heterogeneity criteria lead to experimentally validated gene modules in human glioblastoma [80]. Because of the strong evidence

of correlations between protein co-expression and protein interactions [72, 81], clustering methods can effectively be used for the inference of protein–protein interaction networks. Such methods have, for instance, been applied to the *Saccharomyces cerevisiae* interactome [81]. While it is possible to deduce functional similarities among gene products by clustering the genes according to probabilistic correlations in their micro-array expression profiles [72, 81], clustering is less useful for the inference of specific gene-regulatory relationships, since correlation between mRNA expression profiles associated with two or more genes does not necessarily imply a causal relationship among these genes. Nevertheless, clustering of genes on the basis of expression profiles can give insight into the classes of genes that respond in a similar manner to varying conditions, and that might therefore be co-regulated [72, 80, 82, 83], and these functional clusters can be used as a starting point for determining gene-regulatory relationships.

Data-mining can be used to infer protein–protein interactions, gene-regulatory relationships, and even metabolic pathways [73, 74]. Data-mining schemes typically extract information on relationships between two entities based on the statistical co-occurrence of features of interest associated with the entities, for example, their inclusion in databases and biomedical journals [74]. In this case, by correlating the frequencies of the keywords with the probability that a given interaction is addressed in a paper (estimated from a training set) [73, 84], machine learning algorithms can determine whether or not a particular paper is likely to discuss a specific interaction. Algorithms of this nature have been used extensively to augment the yeast protein–protein interaction network [73]. Search tools such as STRING (http://string.embl.de/) employ similar data-mining methods for the inference of both direct and indirect protein–protein interactions in eukaryotes and prokaryotes [85, 86].

Bayesian network protocols are used to integrate multiple data sets, having variable reliability, for network inference [75, 87]. Bayesian networks are graphical representations of joint probability distributions, consisting of two parts, a directed, acyclic graph (DAG) that qualitatively describes the dependency relationships among variables in the system, and a set of local joint probability distributions that statistically convey these relationships [87]. A node $j$ is assumed to depend conditionally on a node $i$ if there is a directed path from $i$ to $j$ in the DAG. Knowing that the observed state of node $j$ is conditionally dependent on the states of (some or all) other nodes with which $j$ has a dependency relationship, the task of Bayesian inference is to find the posterior distribution of nodes that engender the observed state of node $j$. The posterior distribution is a combined probability, and the distribution yielding the highest logarithmic value (i.e. the highest Bayesian score) is chosen as the best fit to the data.

Bayesian network analysis typically produces multiple candidate networks for a single data set, and these candidates are scored against the observed data and against one another [88]. The links in each network's initial DAG can be established either randomly, or heuristically, based on an initial assessment of the experimental data. An iterative search-and-score algorithm is generally employed: for example, a genetic algorithm might randomly swap edges between two candidate networks, recalculating the posterior distribution and Bayesian score for each network and discarding the networks whose score is lower than a previous score [88].

Benchmarks for weighting Bayesian networks are typically obtained by determining the likelihood, within a given data set, of observing particular events that have been confirmed in the open literature; inferred dependency relationships can then be weighted by the overall accuracy of the data set to which they belong [75]. Bayesian networks have been employed to sort yeast proteins into functional groupings based on a multi-variate Bayesian score achieved by each protein [75]. These methods have also been applied to the inference, from DNA hybridisation-array data, of gene–protein interactions of the *S. cerevisiae* cell cycle [87]. A dynamic extension of the naïve Bayesian network approach can also be used to uncover regulatory relationships from time-course data [87], giving a time-dependent joint probability distribution. However, these dynamic Bayesian networks require extensive pre-existing knowledge of the biological system under consideration, and such knowledge is not always available.

### 3.2 Deterministic methods

Deterministic methods employed for the inference of gene-regulatory networks from time-course gene expression (micro-array) data seek to correlate the rate of change in the expression level (mRNA concentration) of a given gene with the levels of other genes in the network by describing the interdependence in expression in one of two ways: continuous deterministic methods postulate a system of differential equations [89, 90], while Boolean and other logical methods predict a discrete relationship [91, 92]. Deterministic methods based on systems of linear differential equations have, for example, been used to infer gene-regulatory networks in *B. subtilis* [90] as well as regulatory networks specific to the central nervous system of the rat [89]. Experimental data on gene expression levels is substituted into the relational equations, and the ensuing system of equations is then solved for the regulatory relationships between two or more components [90]. Because often there are far more biochemical components in the network than there are experimental time points, multiple networks will be possible solutions; for this reason, a plausible optimisation procedure (e.g. assuming maximum sparseness) is also incorporated into the method [90]. Although differential equation-based deterministic methods tend to be fairly accurate, nonlinear deterministic methods, in particular, are computationally cumbersome and difficult to implement.

Deterministic Boolean methods (and other logical methods), in their simplest form, replace the linear or nonlinear differential equations, with Boolean logical functions for each node [91, 92]. The standard Boolean method assumes that the expression level of each node can be approximated with a binary variable: each node is either expressed/ON or not expressed/OFF. Each node's logical function is found by determining the minimum set of nodes whose (changing) expression levels can explain the observed changes in state of the given node in all experimental trials. Generally, an optimization technique, such as the coefficient of determination [93] is employed for this inference. It is possible that more than one minimum set may be found for a particular node, and, in this case, multiple networks explain the experimental observations. While less accurate than differential equation-based methods, Boolean methods are usually more computationally tractable, and algorithms, such as REVEAL (reverse engineering algorithm) [94], offer promising first steps towards large-scale network inference.

Metabolic pathway reconstruction from known stoichiometric information is usually performed by constraint-based deterministic methods [95]. Particular network states of interest can be found from flux balance analysis [96, 97]; for time series analyses S-systems, power-law approximations

of enzyme-catalysed reactions, are used [98]. Recently, a bi-level linear optimisation strategy that first selects an optimal active subset of a predetermined set of metabolic reactions, and then optimises the metabolic flux distribution, was proposed [99]. A similar framework identified the changes in an *E. coli* genome-scale metabolic model that are needed to minimise the discrepancy between model predictions of optimal flux distributions and experimentally measured flux data [100].

### 3.3 Hybrid methods

The accuracy and realism of network inference methods is increased by incorporating stochastic fluctuations in expression levels due to the potential variability of the synthesis and degradation rates of network constituents [92, 101]. For example, probabilistic Boolean methods [91, 102] bridge the gap between discrete and continuous deterministic methods, while incorporating the effects of uncertainty by assigning to each node $N$ Boolean functions, each with some probability of being chosen to advance the state of the node to which it belongs [91, 102]. An optimisation procedure, usually a machine learning algorithm [93, 102], then selects the updating function for each node at each time point. A novel regulatory network involved in embryonic segmentation and muscle development in *D. melanogaster* [103] was recently produced using a probabilistic Boolean method. Although probabilistic Boolean networks are attractive in that they maintain the large-scale inference ability of standard Boolean methods while relaxing the determinism of the basic method, they can be problematic to use for the extraction of large networks, since successful implementation of the optimisation algorithms requires the estimation of a large number of parameters, and since the amount of training data needed for useful predictions is often prohibitively extensive.

The majority of network inference methods presented in this section use node-level (expression) information to infer causal relationships. There is also a complementary problem: inferring interactions from indirect causal relationships. Indeed, experimental information about the involvement of a protein in a process is often restricted to evidence of differential responses to a stimulus in wild-type organisms versus an organism where the respective protein's expression or activity is disrupted. These observations can be incorporated by two intersecting paths in an (incompletely mapped) interaction network; the inference algorithm must integrate indirect and direct evidence to find a network consistent with all experimental observations [104]. This inference problem is less studied [104–106], but we expect it will play an increasing role when integrating information from disparate data sources.

## 4 Network measures

With a set of experimentally determined or inferred interactions in hand, the relevant cellular network can be constructed. A graph-theoretical analysis of a cellular network can then provide powerful biological insights into the structural organisation and function(s) of the system. Here we present a handful of graph-theoretical measures that can serve as the basis of such an analysis.

### 4.1 Degree and degree distribution

The degree of a node is the number of edges adjacent to that node. If the directionality of interaction is important, a node's total degree can be broken into an in-degree and an out-degree, quantifying the number of incoming and outgoing edges adjacent to the node. For example, in Fig. 4, node $J$ has both in-degree and out-degree 2. In a graph whose edges are quantified by weights one can also define a node strength, the sum of the weights of the edges adjacent to the node. For example, in Fig. 4, the strength of the node $P$ is 4. While the degree of a specific node is a local topological measure, this local information can be synthesised into a global description of the network by reporting the degrees of all nodes in the network in terms of a degree distribution, $P(k)$ [$P(k_{\text{in}})$ and $P(k_{\text{out}})$ in directed networks], which gives the probability that a randomly selected node will have degree $k$ (Fig. 4). The degree distribution is obtained by first counting the number of nodes, $N(k)$, with $k = 1, 2, 3, \ldots$ edges, and then dividing this number by the total number of nodes, $N$, in the network (the same procedure can be employed to find in- and out-degree distributions for a given directed network).

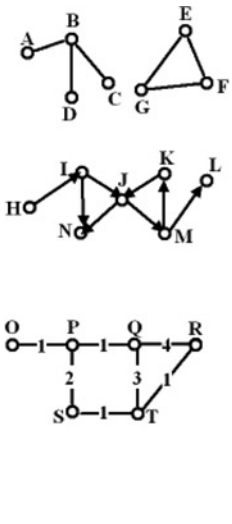
The majority of cellular networks has been shown to have (out-) degree distributions that are scale-free, meaning that the high diversity of node degrees precludes the existence of a typical node that could be used to characterise the rest of the nodes in the network (reviewed in [107]). The degree distribution of these scale-free networks is close to a power-law: $P(k) \simeq Ak^{-\gamma}$, where $A$ is a normalisation constant, and where the degree exponent $\gamma$ is typically similar for similar networks. The degree distributions of protein interaction networks, metabolic networks and the out-degree distribution of most gene-regulatory networks, for example, are power laws with $2 < \gamma < 3$ [4, 108, 109].

### 4.2 Clustering coefficient

The clustering coefficient quantifies the extent to which a node's first neighbourhood is a completely connected subgraph (clique) [110]. Mathematically, the local clustering coefficient is given by

$$C_i = \frac{2E_i}{k_i(k_i - 1)}$$

where $E_i$ is the number of edges connecting the immediate neighbours of node $i$, and $k_i$ is the degree of node $i$. For example, in Fig. 4, the clustering coefficient of node $E$ is 1, since nodes $E$, $F$, and $G$ are maximally cohesive (i.e. form a clique). By averaging the clustering coefficients of all nodes in a network to obtain an average clustering




| $k$ | $P(k)$ | $C(k)$ |
|---|---|---|
| 1 | 3/7 | - |
| 2 | 3/7 | 1 |
| 3 | 1/7 | 0 |

| $k_{in}$ | $P(k_{in})$ | $k_{out}$ | $P(k_{out})$ |
|---|---|---|---|
| 0 | 1/7 | 0 | 2/7 |
| 1 | 4/7 | 1 | 2/7 |
| 2 | 2/7 | 2 | 3/7 |

| *Node* | *Strength* | *Betweenness (weighted)* |
|---|---|---|
| O | 1 | 0 |
| P | 4 | 5 |
| Q | 8 | 0 |
| R | 5 | 0 |
| S | 3 | 3.5 |
| T | 5 | 4 |

**Fig. 4** *Illustration of frequently used network measures*

coefficient, an idea of the strength of the connectivity within the network can be established. Protein–protein interaction networks and metabolic networks [8] exhibit large average clustering coefficients, indicating a high level of redundancy and cohesiveness. Alternatively, the average clustering coefficient of nodes with degree $k$ can be plotted as a function of node degree, $C(k)$ (Fig. 4). It has been found that for a wide variety of cellular networks, this clustering-degree relation has the functional form $C(k) = B/k^{\beta}$, where the exponent $\beta$ typically falls between 1 and 2 [111, 112].

### 4.3 Connectivity, paths, distances, efficiency and graph components

Two nodes of a graph are connected if a sequence of adjacent nodes, a path, links them [113]. A path can signify a transformation route from a nutrient to an end product in a metabolic network, or it could represent a chain of ligand-induced reactions from the source to the sink in a signal transduction network. The distance (path length) between any two nodes in a network is defined to be the number of edges in the shortest path connecting those nodes. For example, in Fig. 4, the distance between nodes $V$ and $Z$ is 1 and the distance between nodes $A$ and $C$ is 2 (along the ABC path). If the edges of a network are weighted (e.g. with rate constants), then the distance between two nodes will be the sum of the edge weights along the path for which this sum is a minimum [114]. Often, the average path length, $d = \langle d_{ij} \rangle$ that is, the average number of edges in the shortest path between any two nodes in a network (Fig. 4), scales with the natural logarithm of the number of nodes in the graph: $d \sim ln(N)$. This small world [110] property implies that path lengths remain small, even if the networks become very large, and it is a feature of metabolic and protein interaction networks.

Particularly when a network is directed, it is possible that by starting at an edge adjacent to a given node and tracing a path along consecutive edges, only a fraction of the nodes in the network will be accessible to the starting node. If a network is either directed or unconnected, it is often more advantageous to define the graph's global efficiency $= \langle 1/d_{ij} \rangle$ [115, 116]. Unconnected nodes' distance is infinite by definition, and thus these node pairs do not contribute to the network's efficiency. If, however, a path does exist between every pair of nodes in a network, the network is said to be connected. Directed networks having directed paths, in both directions, between every pair of nodes are said to be strongly connected. Even if a cellular network is not (strongly) connected it is beneficial to identify connected partitions of the network. For example, a directed network has one or several strongly connected components, a subgraph(s) whose node pairs are connected in both directions. Each strongly connected component is associated with an in-component (nodes that can reach the strongly connected component, but that cannot be reached from it) and an out-component (the converse). For example, in Fig. 4, nodes $J$, $K$, and $M$ of the directed graph constitute a strongly-connected component of the graph. The in-component of the graph contains nodes $H$ and $I$, while the out-component contains nodes $N$ and $L$. It has recently been suggested that the nodes of each of these components share a component-specific task within a given network. In signal transduction networks, for example, the nodes of the in-component tend to be involved in ligand-receptor binding; the nodes of the strongly connected component form a central signaling subnetwork; and the nodes of the out-component are responsible both for the transcription of target genes as well as for phenotypic changes [117, 118]. By identifying large connectivity classes within a network, one may be able to gain a sense of how the network is organised functionally.

### 4.4 Betweenness centrality, sources and sinks

The number, (net) directionality, and strength of connections associated with a given node are measures of that node's local centrality, and can be synthesised into distributions over all nodes in a network to give valuable insight both into the heterogeneity of node interactivity levels within a cellular network, and into the flow of information, mass, or other entities through the network. In particular, the sources and sinks of the network – those nodes with only outgoing or incoming edges, respectively, represent the initial and terminal points of the flow. In signal transduction networks, for example, extracellular ligands and/or their receptors are typically sources, acting as the injection points for chemical signals; these chemical signals then terminate at effectors, the networks' sinks [117]. If a node is neither a source nor a sink, its betweenness centrality the number of (shortest) paths from node $s$ to node $t$ passing through the node, divided by the total number of (shortest) $st$-paths (Fig. 4) indicates the importance of that node to the propagation of flow within the network [119, 120]. For example, in Fig. 4, the weighted and unweighted betweenness centralities of node $P$ are higher than the betweenness centralities of the other nodes in the graph, since all paths involving node $O$ must pass through node $P$. One can similarly define the betweenness centrality of an edge [121]. Even though a node's betweenness centrality is not necessarily correlated with its degree, betweenness centralities are usually power-law distributed, with a characteristic exponent close to 2 [122]. Holme *et al.* have demonstrated that while the most ubiquitous substrates in biochemical pathways may not have the highest degrees in the network, they often have the highest betweenness centralities [123].

## 5 Network properties of cellular networks

To date, few cellular networks have been reconstructed and analysed in full. However, transcriptional regulatory maps exist for *E. coli* [124] and *S. cerevisiae* [4, 108, 125], and protein–protein interaction maps have been constructed for a variety of organisms, including viruses [126], prokaryotes such as *Helicobacter pylori* [127], and eukaryotes such as *S. cerevisiae* [128], *C. elegans* [46], *D. melanogaster* [45] and *H. sapiens* [47]. Graph-theoretical analysis of these and other cellular networks has yielded a wealth of information regarding structural organisation and functioning at the cellular level.

Both prokaryotic and eukaryotic transcription networks exhibit an approximately scale-free out-degree distribution, signifying the potential of transcription factors to regulate a multitude of target genes. The in-degree distribution is a more restricted exponential function, illustrating that combinatorial regulation by several transcription factors is observed less than regulation of several targets by the same transcription factor (see Fig. 5). Neither the *E. coli* nor the yeast transcription network has strongly connected components, suggesting a uni-directional, feed-forward type regulation mode. Note that although purely transcriptional feedback loops are under-represented, feedback is still frequently accomplished through the protein-level regulation of transcription factors, often via auto-regulation. The subgraphs found by following the paths that start from

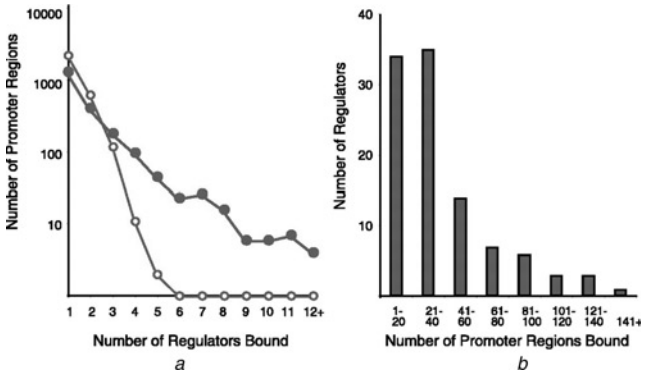



**Fig. 5** *Genome-wide distribution of transcriptional regulators in S. cerevisiae*

*a* Full symbols represent the number of transcription factors bound per promoter region (corresponding to in-degree of regulated gene)
Open symbols represent in-degree distribution of a comparable randomised network.
*b* Distribution of number of promoter regions bound per regulator (i.e. out-degree distribution of transcription factors)
Reproduced from [4] with permission from the American Association for the Advancement of Science

non-transcriptionally regulated genes have relatively little overlap [129], reflecting the fact that distinct environmental signals tend to initiate distinct transcriptional responses. The source–sink distances are small in both networks, and the longest regulatory chain has only four (in *E. coli*) or five (in *S. cerevisiae*) edges (see Fig. 5).

The current versions of protein interaction maps are, by necessity, incomplete, and also suffer from a high rate of false positives. Despite these drawbacks, there is an emerging consensus in the topological features of the maps of different organisms (Fig. 6). For example, all protein interaction networks have a giant connected component and the distances on this component are close to the small-world limit given by random graphs [45, 112]. This finding suggests an illustration of pleiotropy, since perturbations of a single gene or protein can propagate through the network, and have seemingly unrelated effects. The degree distribution of the yeast protein interaction network is approximately scale-free (see Fig. 6). The *Drosophila* protein network exhibits a lower-than-expected fraction of proteins with more than 50 interacting partners. This deviation is suspected to be caused by incomplete coverage and could change as more interactions are discovered, as was the case for the yeast protein interaction network [45, 112, 130]. The heterogeneous clustering-degree function $C(k) = B/k^{\beta}$ [112], where the exponent $\beta$ is around 2, and the inverse correlation between the degree of two interacting proteins [131] indicate that the neighbourhood of highly connected proteins tends to be sparser than the neighbourhood of less-connected proteins.

The largest reconstructed signal transduction network was synthesised from more than 1200 articles in the experimental literature and contains 1259 interactions among 545 cellular components of the hippocampal CA1 neuron [118]. This network exhibits an impressive interconnectivity: its strongly-connected component (the central signalling network) includes 60% of the nodes, and the subgraphs that start from various ligand-occupied receptors reach most of the network within 15 steps. The average input–output path-length is near 4, suggesting the possibility of very rapid response to signalling inputs. Both the in- and out-degree distributions of this network are consistent with a power-law with an exponent around 2, and the highest degree nodes include the four major protein kinases (MAPK, CaMKII, PKA and PKC).

All metabolic network representations indicate an approximately scale-free [7, 8, 132] or at least broad-tailed [133] metabolite degree distribution. Fig. 7 illustrates the idea that functionally different metabolities tend to cover different ranges of the degree spectrum. The degree distribution of enzymes is strongly peaked, indicating that enzymes catalysing several reactions are rare [7]. The variability of metabolite degrees can be accounted for if they are functionally separated into high-degree carriers and low-degree metabolites unique to separate reaction modules (such as catabolism or amino acid biosynthesis) [132]. However, such a picture does not seem to explain the frequency of intermediate degrees. The clustering-degree function follows the relationship. $C(k) \simeq 1/k$.

The substrate and reaction graphs indicate a remarkably small and organism-independent average distance among metabolites and reactions [7, 8]. If the preferred directionality of the reactions is known and is taken into account, only the largest strongly connected component (whose nodes can

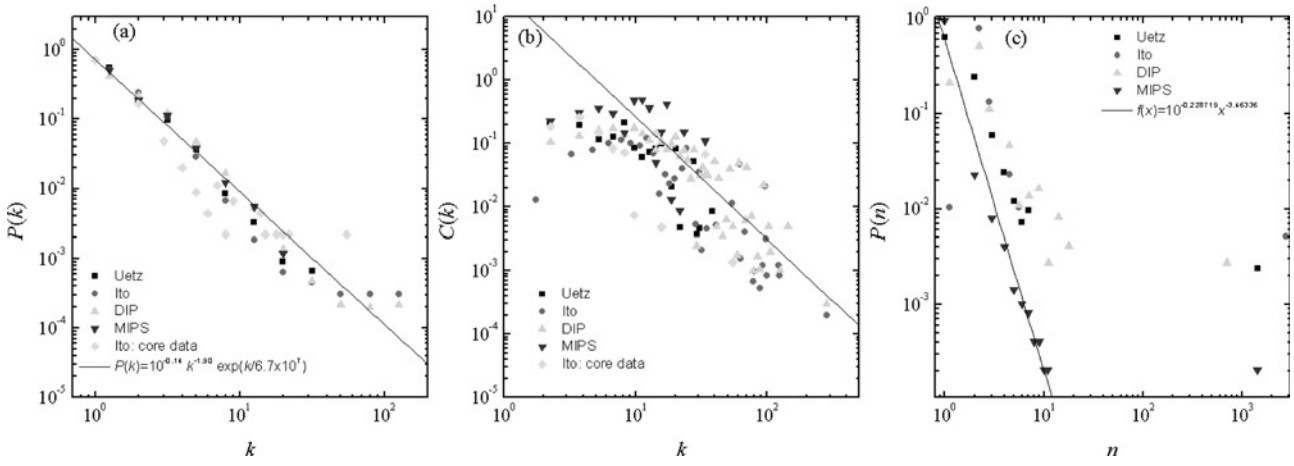



**Fig. 6** *Topological properties of the yeast protein interaction network constructed from four different databases*

*a* Degree distribution, solid line corresponds to a power law with exponent $\gamma = 2.5$
*b* Clustering coefficient – degree function, solid line corresponds to function $\boldsymbol{C(k) = B/k^2}$
*c* Size distribution of connected components
All networks have a giant connected component of more than 1000 nodes (on the right) and a number of small isolated clusters
Reproduced from [111] with permission from Wiley InterScience

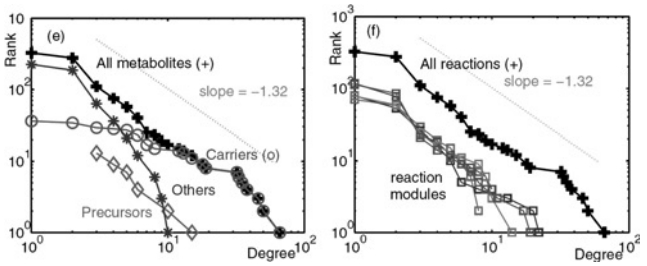



**Fig. 7** *Rank (cumulative distribution) of metabolite node degree (left panel) and reaction node degree (right panel) for metabolic networks of H. pylori*

Straight lines correspond to a power-law degree distribution with exponent $\gamma = \text{slope} + 1 = 2.32$
Reproduced from [131] with permission from the American Physical Society

reach each other in both directions) has a well defined average path length. While this average path length is still small in all the organisms studied, the strongly connected component itself contains less than 50% of the nodes [134]. An alternative representation of the *E. coli* metabolic network defines the edges among metabolites as structural changes that convert the source metabolite into the target metabolite [133]. As separate reactions can involve the same structural change in a metabolite, this alternative representation has less than half as many edges as the metabolite graph defined by [7], and consequently it yields metabolite distances that are twice as high, on average.

The general architectural features of molecular interaction networks described so far are shared to a large degree by other complex systems, ranging from technological networks to social networks. While this universality is intriguing, it is arguably more important to discern whether, and how, the graph properties of cellular networks reflect their functional and evolutionary constraints.

### 5.1 Hubs

In a scale-free network small-degree nodes are the most abundant, but the frequency of high-degree nodes decreases relatively slowly. Thus, nodes that have degrees much higher than average, so-called hubs, exist. Because of the heterogeneity of scale-free networks, random node disruptions do not lead to a major loss of connectivity, but the loss of the hubs causes the breakdown of the network into isolated clusters [107]. The validity of these general conclusions for cellular networks can be verified by correlating the severity of a gene knockout with the number of interactions the gene's products participate in. Indeed, as much as 73% of the *S. cerevisiae* genes are non-essential – i.e. the knockout has no phenotypic effects [135]. This confirms the cellular networks' robustness in the face of random disruptions. The likelihood that a gene is essential (lethal) or toxicity-modulating (toxin-sensitive) correlates with the number of interactions its protein product has [130, 136]. This indicates that the cell is vulnerable to the loss of highly interactive hubs. It should be noted that different network representations can lead to distinct sets of hubs and there is no rigid boundary between hub and non-hub genes or proteins. Among the most well known examples of hub proteins is the tumor suppressor protein p53, which has an abundance of incoming edges – interactions regulating its conformational state (and thus its activity) and its rate of proteolytic degradation – and which also has many outgoing edges in the genes whose transcription it activates. The tumor suppressor p53 is inactivated by a mutation in its gene in 50% of human tumors, corroborating the fact that cellular networks are vulnerable to loss of their most connected hubs [137]. High interactivity is not the only marker of functional importance, however: low-degree nodes in genome-wide metabolic networks of various micro-organisms are almost as likely to be critical to the overall network functions as high-degree nodes [138].

### 5.2 Modularity

Cellular networks have long been thought to be modular, composed of functionally-separable sub-networks corresponding to specific biological functions [139]. Since genome-wide interaction networks are highly connected, modules should not be understood as disconnected components but rather as components that have dense intra-component connectivity but sparse inter-component connectivity. Several methods have been proposed to identify functional modules on the basis of the physical location or function of network components [140], or on the topology of the interaction network [141, 142]. The challenge is that modularity does not always mean clear-cut subnetworks linked in well defined ways, since there is a high degree of overlap and cross-talk between modules [143]. A heterogeneous degree distribution, inverse correlation between degree and clustering coefficient (as seen in metabolic and protein interaction networks) and modularity, taken together, suggest hierarchical modularity, in which modules are made up of smaller and more cohesive modules, which themselves are made up of smaller and more cohesive modules etc. [144].

### 5.3 Motifs and cliques

Cellular networks contain recurring interaction motifs, small subgraphs that have well defined topologies (Fig. 8). Interaction motifs such as autoregulation (usually a negative feedback) and feed-forward loops have a higher abundance in transcriptional regulatory networks than expected from randomly connected graphs with the same degree distribution [124, 129]. Protein interaction motifs such as short cycles and small, completely connected subgraphs are both abundant [45] and evolutionarily conserved [145], partly because of their enrichment in protein complexes. Feedforward loops and triangles of scaffolding (protein) interactions are also abundant in signal transduction networks, which additionally contain a significant number of feedback loops, both positive and negative [118]. Yeger-Lotem *et al.* have identified frequent composite transcription/protein interaction motifs such as interacting transcription factors coregulating a gene or interacting proteins being coregulated by the same transcription factor [146]. The abundant motifs of integrated mRNA/protein networks are often signatures of higher-order network structures that correspond to biological phenomena such as interacting transcription factors regulating the same target gene or co-regulation of members of a protein complex [147].

The functional relevance of recurring network motifs has been investigated both theoretically and experimentally. Coherent feed-forward loops have been shown to cause sign-sensitive delay: coherent feed-forward loops whose output gene is regulated by an AND function filter out brief signal fluctuations [148], while coherent feed-forward loops with an OR or SUM function prolong expression following deactivation of the signal [149]. Incoherent feed-forward loops, having an AND NOT regulation of the output gene, accelerate signalling through an overshoot

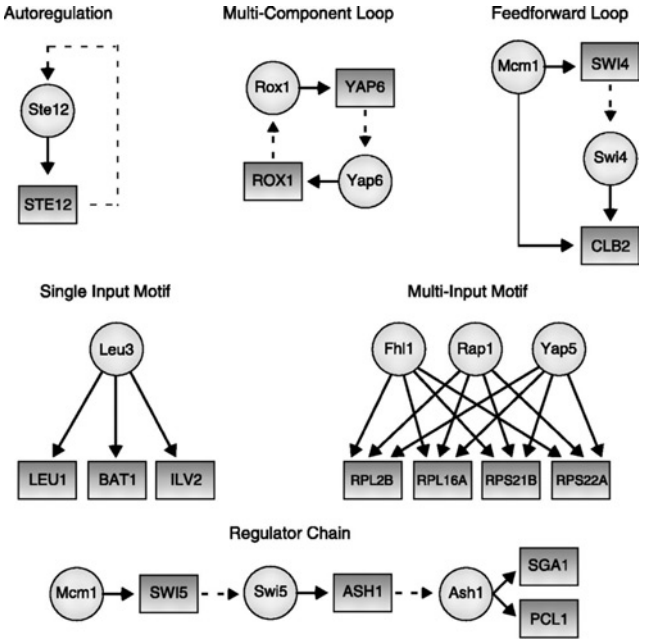



**Fig. 8** *Examples of network motifs in the yeast transcriptional regulatory network*

Regulators are represented by circles
Target gene promoters are represented by rectangles
Binding of a regulator to a promoter is indicated by solid arrows
Genes encoding regulators are linked to their respective regulators by dashed arrows
Reprinted from [4] with permission from the American Association for the Advancement of Science

dynamics [149]. The comparative abundance of negative feedback loops in the early steps of signal transduction networks and of positive feedback loops at later steps suggest that weak or short-lived signals are filtered by early barriers posed by negative feedback loops, while strong and persistent signals are amplified and are able to evoke a biological response [118].

### 5.4 Path redundancy

Any response to a perturbation requires that information about the perturbation spreads within the network. Thus the short path lengths of metabolic, protein interaction and signal transduction networks (their small world property) [7, 118, 130] is a very important feature that ensures fast and efficient reaction to perturbations. Another very important global property related to paths is path redundancy, or the availability of multiple paths between a pair of nodes [150]. Either in the case of multiple flows from input to output, or contingencies in the case of perturbations in the preferred pathway, path redundancy enables the robust functioning of cellular networks by relying less on individual pathways and mediators.

## 6 Graph models

An often-used means of testing how organizational features such as those described in the previous section reflect the networks' functions is to construct model networks based on features or assembly principles deemed most salient. The predictive power of the model is determined by a graph-theoretical comparison of the model network to the original. In the following we examine three general families of network models as well as a family of network models formulated specifically for intracellular networks.

### 6.1 Erdős–Rényi (ER) random graphs

Mid-twentieth-century work on uniformly random graphs [151] pioneered many of the basic measures and analytic techniques mentioned in the previous sections. An Erdős–Rényi (ER) random graph is formed by randomly connecting $N$ nodes with $E$ edges. For large $N$, the degree distribution of such a graph is Poissonian, implying that most nodes have degree $k$, close to the average degree in the graph, $\langle k \rangle = 2E/N$. Therefore, unlike most cellular networks, ER random graphs are largely homogeneous in degree. In addition, the average clustering coefficient of ER random graphs scales inversely with the size of the network, such that $\langle C \rangle = \langle k \rangle / N$, and the clustering coefficient distribution of an ER random graph (unlike that of most cellular networks) is independent of node degree, peaking at a value equal to the connection probability $p$. Finally, the average path length of ER random graphs $\langle d \rangle \simeq ln\ (N)/ln\langle k \rangle$, remains small even for large networks [113], and is consistent with the average path length of several real networks [107].

### 6.2 Scale-free random graphs

Scale-free random graphs adhere to a prescribed degree distribution, though individual links in the graphs are established randomly [152]. Scale-free random graphs have smaller average path lengths [153] than comparably sized ER random graphs, and they exhibit clustering coefficients similar to their ER counterparts [154]. Thus, while the clustering coefficient of biological networks is not captured by scale-free random graphs, scale-free random graphs are approximately comparable to scrambled, but degree distribution-preserving versions of real networks. For this reason, they serve as a better null model of biological networks than do ER random graphs, a point corroborated by the fact that scale-free random graphs are often used as baselines from which to establish statistical significance thresholds for features and properties of biological networks [155, 156]. While scale-free random graphs are, by definition, able to capture the scale-free degree distribution found in most real networks, they make no attempt to explain why such heterogeneity in connectivity exists in real networks. Accounting for this feature necessitates a shift from modelling network topology to modelling network assembly.

### 6.3 Evolving network models

A large (and growing) class of network models addresses the question of how scale-free topologies arise in real networks by describing network assembly and evolution. The simplest of these evolving network models is the Barabási–Albert (BA) model [157], which introduced two core assumptions: growth (i.e. an increase in the number of nodes and edges over time) and preferential attachment (i.e. a greater likelihood that nodes of high degree will acquire new edges). The BA model assumes linear growth and proportional preferential attachment, and leads to a power-law degree distribution $P(k) = Ak^{-3}$ that captures the upper end of the range of observed degree-distribution exponents in biological networks. Networks generated with the BA algorithm have small average clustering coefficients, a constant clustering-degree function $C(k)$ [111] and slightly smaller average path lengths than are found in comparable random graphs [158], features that prevent them from capturing all the topological characteristics of real networks. Numerous models, augmenting linear growth and

proportional preferential attachment with features such as: nonlinear attachment [159], initial attractiveness of isolated nodes [160], accelerated growth [161], aging [162], fitness [163] and node loss [164] have offered successful solutions for the shortcomings of the basic BA model. For example, the asymptotically linear preferential attachment [160, 164] of the linear preferential attachment (LPA) class of models enables tuning of the degree exponent; for recent reviews of evolving network models see [165, 166].

A recent model proposed by Ravasz *et al.* is based on a self-similar growth pattern and not on preferential attachment [144]: here, the network grows by iterative network duplication and subsequent integration of the duplicated elements to the network's original core. The net result of the replication model is a degree-distribution exponent,

$$\gamma = 1 + \frac{\log(n)}{\log(n-1)}$$

where $n$ is the size of the seed graph. For small $n$, the replication model produces a degree-distribution exponent very close to 2, comparable to what is seen in cellular networks. In addition, in contrast to all previous models, the replication model produces a clustering coefficient that is independent of the size of the network and that scales inversely with node degree, properties that also seem to characterise protein interaction and metabolic networks.

### 6.4 Models of cellular network evolution

The topology of cellular networks is shaped by dynamical processes such as gene duplication or point mutations, occurring on evolutionary time scales. Interestingly, both gene duplications and point mutations, unique biological processes, lead to a preferential increase of the degree of highly connected proteins, that is to preferential attachment [167, 168]. Estimates of gene duplication rate and the rate at which point mutations lead to the gain or loss of protein interactions indicate that the latter is two orders of magnitude higher [169]. Duplication events often lead to asymmetric change in edge dynamics for the two gene copies; moreover a large fraction of duplicate genes (44.4% in yeast [170]) are deleted during evolution, making it difficult to trace ancient gene duplication events. Several growing network models based on random gene duplication or mutation and subsequent functional divergence have displayed good agreement with the topology of protein interaction networks [169, 171–173]. Note that these network models aim to identify the main evolutionary mechanisms shaping the topology of cellular interaction networks across organisms and not to predict individual gene duplication events. The modelling of the evolution of transcriptional, metabolic and signal transduction networks has added challenges due both to these networks' directed nature and to the complexity of the regulatory mechanisms involved, but rapid progress is expected in these fields [174, 175].

## 7 Evolutionary relevance of the observed topology

The previous section illustrated that generic models based on gene duplication attain a reasonable agreement with the large-scale organisation of cellular networks. More specific lessons can be learned from integrating comparative genomics (identifying substitutions, insertions, fusions or deletions in DNA sequences) and comparative network analyses. Overlaying networks of orthologous proteins across a large number of genomes revealed conserved functional modules of physically interacting, biochemically related, or genetically interacting (compensatory) proteins [176]. Such analysis also enables prediction of new functionally relevant interactions, recognition of conserved co-ordination between cellular processes, and evaluation of the effects of loss of duplicated or mutated genes [177]. In this section we will briefly discuss the evolutionary relevance of node degree and network motifs.

Heterogeneous networks are vulnerable to targeted attack of the highly connected nodes (hubs) described in the previous section, while having a significant tolerance to the loss of other nodes. Thus one can hypothesise that hubs are subject to severe selective and evolutionary constraints. Hahn *et al.* [178] have correlated the rate of evolution of yeast proteins with their degree in the protein interaction network, and the rate of evolution of *E. coli* enzymes with their degree in the core metabolic reaction graph constructed by [8]. Although they obtained statistically significant (albeit weak) negative correlation between yeast protein degree and evolution rate, no such correlation was evident in the *E. coli* enzyme network. The latter result has the caveat that the edges linking enzymes do not correspond to interactions; thus further studies are needed to gain a definitive answer.

The abundant transcription factor motifs of *E. coli* and *S. cerevisiae* do not show common ancestry but are a result of repeated convergent evolution [179]. These studies, taken together with the dynamical repertoire of the interaction motifs, suggest that these motifs represent elements of optimal circuit design [13, 180]. Vergassola *et al.* observed cooperative co-evolution within cliques of interacting proteins of *S. cerevisiae* [181], implying that co-operative compensatory mutations are a globally relevant mechanism to preserve the specificity in the assembly of complexes throughout evolutionary divergence processes. Cliques of interacting proteins are simple instances of motifs, suggesting that the multi-point coevolution correlations might be a general feature of the modular architecture of biological networks.

A recent study inferring the conserved patterns of transcriptional regulatory networks of 175 prokaryotes based on the known *E. coli* transcription network found that prokaryotic transcription factors are typically less conserved than, and evolve independently of, their target genes [182]. As illustrated in Fig. 9, the study suggests that organisms with different lifestyles and environments have convergently acquired similar network structures approximating a scale-free topology. Metagenomics analysis of microbial communities also found environment-specific genes in closely related organisms [183]; thus it will be interesting to combine such comparative analyses with growth models addressing the evolution of higher organisms.

## 8 Modelling the dynamics of cellular networks

The graphs comprising cellular networks are static representations of dynamic processes. Moreover, the nodes of cellular networks often represent entire populations of proteins or molecules without reflecting the abundance or activity of these populations. To capture the changes in gene expression levels or in protein/molecule abundances in response to external or internal stimuli, the interaction networks must be augmented with node-level information characterising the expression, abundance, and activity, that is the state, of each node. A dynamical model of an interacting system is based on the premise of locality in

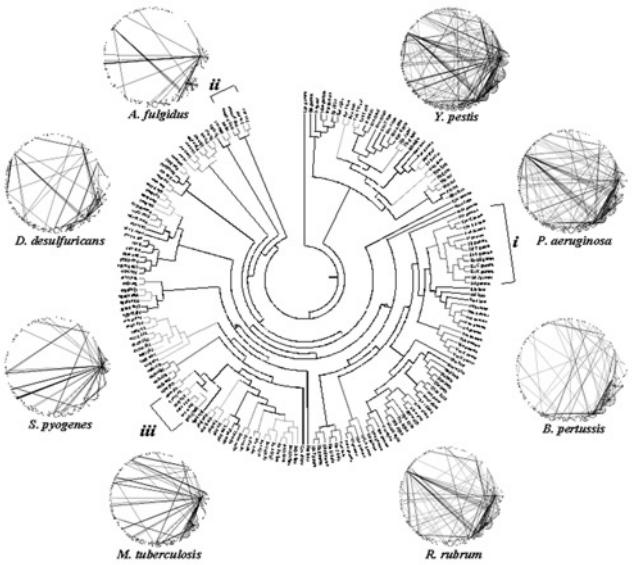



**Fig. 9** *Proposed distance tree of prokaryotic transcriptional regulatory networks, prokaryotic genomes are clustered according to the interactions they conserve*

(i) Genomes in the same phylogenetic group generally cluster together
(ii) Parasitic genomes cluster together
(iii) Genomes of organisms with a similar lifestyle but belonging to different phylogenetic groups are clustered together
Reprinted from [182] with permission from Elsevier

the network space and consists of a set of equations indicating how the state of each node changes in response to changes in the state of its regulators (including itself), where the identity of the regulators is given by the interaction network.

Dynamic models have as input information: (i), the interactions and regulatory relationships among components (i.e. the interaction network); (ii), the manner in which the strength of the interactions depends on the state of the interacting components (i.e. the transfer functions, including kinetic parameters for each interaction), and (iii), the initial state of each component in the system. Given these inputs, the model will output the time evolution of the state of the system, for example the system's response to the presence or absence of a given signal. Due to the demanding prerequisites of dynamic modelling, most dynamic networks so far constructed have been quite small (e.g. [15, 184–186]). In this section we briefly outline some of the most promising techniques so far developed for modelling the dynamics of cellular networks, and we review the successes and implications of these models' results.

### 8.1 Continuous and deterministic models

These are formulated as differential equations based on mass-action (or more general) kinetics for the production and decay of all components [98]. With sufficiently thorough knowledge of the biochemical interactions comprising a system (i.e. a compilation of all pairwise interactions among the system's components, measurements or estimates of kinetic parameters such as dissociation constants and half-lives, and a known initial state for the system) it is possible to quite accurately reproduce the dynamic behaviour of a complex biological system by describing the constituent interactions as coupled (usually, nonlinear) differential equations. For example, using a continuous deterministic model, von Dassow *et al.* reproduced the expression patterns of segment polarity genes in *D. melanogaster*, and demonstrated that these patterns are remarkably robust to changes in the kinetic parameters governing the biochemical reactions that result in gene expression [186].

There are a number of approaches aimed at incorporating stochasticity and discrete events into dynamic models of cellular systems (for a review, see [187]). In general, stochastic models either formulate a master equation that follows the time evolution of the probabilities of each of a system's possible configurational states (the probability density function), or they append stochastic (noise) terms to differential equations. The former method usually uses a Monte Carlo (Gillespie) algorithm to select a state of the system, compatible with the master equation, at each time step. The sequence of these states will form one time course for the system, and the results of multiple iterations will then be interpreted statistically. The time-evolution of a system's probability density function can also be estimated from stochastic differential equations by using a Monte Carlo algorithm to select (multiple times) the noise addenda to the differential equations; statistics are then used to interpret the results in terms of probability densities [188, 189]. While stochastic master equation methods are tractable for small systems, their computational complexity becomes prohibitive as the size of the system grows, and it is often necessary to work with the states of classes of interacting entities, instead of with individual entities, a simplification that may be too restrictive for cellular systems in which the individual entities of a single class could be in different states at a given time. Recent algorithms such as StochSim (http://www.pdn.cam.ac.uk/groups/comp-cell/StochSim.html) discretise the chemical master equation by finding the probability of pairwise interaction between individual entities in the system on the basis of experimentally determined reaction rates; the state of the entire system is then manifest from the states of its components. StochSim has been applied to signalling networks in bacterial chemotaxis [188, 190] and has been shown to be asymptotically equivalent to the stochastic master equation approach [189].

### 8.2 Boolean models

These assume binary states for network nodes and are formulated as logical rules describing the change in state of each regulated node as a function of its regulators. The utility of Boolean dynamic models lies in their ability to predict dynamic trends in the absence of detailed kinetic parameters. For example, Boolean models successfully described the wild type or mutant expression patterns of segment polarity genes in *D. melanogaster* and predicted a significant error correcting ability for the gene regulatory network [184, 191]. Boolean modelling also reproduced observed gene expression patterns and mutant behaviour of the floral organs of *Arabidopsis thaliana* [192]. Recently, asynchronous Boolean modelling of the abscisic acid signal transduction network of *Arabidopsis thaliana* led to experimentally verified predictions regarding stomatal responses in wild type plants as well as in plants subjected to gene disruptions and pharmacological interventions [104]. The fact that simple Boolean methods can capture the same (broad) dynamic trends and predict the same asymptotic, biologically viable behaviours as their far more detailed differential equation-based counterparts often makes them a more feasible approach for studying a system's dynamics.

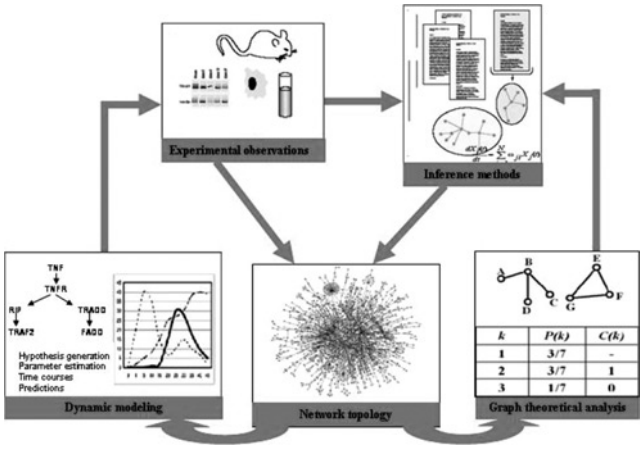



**Fig. 10** *Overview of applications of network analysis in studying cellular networks*

### 8.3 Hybrid dynamic models

These meld Boolean logical functions with continuous synthesis and decay [193–196]. In this approach, the rates of change in concentration of the effector molecules are expressed as differential equations, and combinatorial regulation is described by Boolean functions. Genes that are known to have a clear activation threshold within a narrow range of effector concentrations are assumed to act as threshold-dependent ON/OFF switches, while the expression levels of genes for which activation can occur over a much broader range of effector concentrations are instead described as continuous, graded functions [193, 194]. For example, the dynamics of the cis-regulatory system governing the embryonic expression of the Endo16 sea urchin gene has been successfully modelled with a hybrid model, demonstrating that combinatorial transcriptional regulation is essentially logical and hard-wired into an organism's DNA [194]. This type of hybrid approach is appealing in that its continuous features can incorporate a great deal of quantitative detail while its discrete features allow potential uncertainty in interactions [195]. Being less computationally intensive than strict deterministic models, and requiring fewer initial estimates and less initial input, hybrid approaches can be applied to larger networks than is possible with pure continuous models.

## 9 Conclusions

Our understanding of cellular processes at the systems level grows as the result of an ongoing dialogue and feedback among experimental, computational and theoretical approaches. Fig. 10 shows that high-throughput experiments, in conjunction with inference methods allow optimal network construction. Graph-theoretical analysis of these networks then enables general insight into the topological and functional organisation of cellular regulation and into the evolutionary roots of this organisation. Finally, comparative network analysis feeds back to network inference [79, 80, 197, 198], also expanding the tools of graph theory to incorporate the diversity of molecular interactions.

Most inference methods operate on the assumption that a system can be reduced to interactions taking place on a single temporal scale. However, as experimental data continues to amass, it is becoming increasingly evident both that biochemical reactions occurring within or between cells take place on time-scales that can differ from one another by many orders of magnitude [16], and that these time-scales can be dictated by any number of factors, including the spatial scale on which interactions take place, the types of biomolecules or complexes that are interacting, and the environmental conditions to which the system is subjected [143, 199]. Dynamic models integrate interaction (topological), state (expression or activity) and temporal information regarding a system into a description of its predicted dynamical trajectory in normal and perturbed conditions. Experimental testing of these predictions validates the models' initial hypotheses or generates alternate hypotheses, in either case leading to new biological discoveries.

While the use of extensive dynamic modelling is limited by computational complexity and the availability of detailed transfer functions and kinetic parameters, intermediary approaches such as the work of Ma'ayan *et al.* [118] demonstrate that it is possible to glean a pseudo-dynamical view of cellular processes from graphical representations of cellular networks by mapping the propagation of a chemical signal through the network from an extracellular signal source to sinks within the cell. Similarly, augmenting the currently available directionless interactome networks with information regarding the sources (signals) and outputs of the network and the cause-and-effect (directional) relationships along the edges will significantly enhance their functional information content.

While genome-level interaction maps help us in understanding regulatory design features and evolutionary rules, dynamic modelling of systems with a less than genome-wide scope and specified inputs and outputs allows the identification of key regulatory components or parameters (as opposed to general trends). Emergent qualitative and hybrid modelling techniques give hope that even in the event that exhaustive knowledge of parameters is unreachable, predictive modelling of (sub)cellular processes will still be possible. As experimental and computational techniques continue to improve, adding to the scope and detail captured by network models, future generations of dynamic network models promise to greatly augment our current understanding of biological systems, perhaps causing us to modify our initial assumptions about cellular network structure and function, and unquestionably furthering our understanding of the machinery of the cell.

## 10 References

1 Bollobás, B.: 'Graph theory: an introductory course' (Springer-Verlag, 1979)
2 Pandey, A., and Mann, M.: 'Proteomics to study genes and genomes', *Nature*, 2000, **405**, pp. 837–846
3 Caron, H., van Schaik, B., van der Mee, M., Baas, F., and Riggins, G. *et al.*: 'The human transcriptome map: clustering of highly expressed genes in chromosomal domains', *Science*, 2001, **291**, pp. 1289–1292
4 Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., and Bar-Joseph, Z., *et al.*: 'Transcriptional regulatory networks in Saccharomyces cerevisiae', *Science*, 2002, **298**, pp. 799–804
5 Feinberg, M.: 'Chemical oscillations, multiple equilibria, and reaction network structure', in Stewart, W., Rey, W., and Conley, C. (Eds.): Dynamics of reactive systems (Academic Press, 1980), pp. 59–130
6 Lemke, N., Heredia, F., Barcellos, C.K., Dos Reis, A.N., and Mombach, J.C.: 'Essentiality and damage in metabolic networks', *Bioinformatics*, 2004, **20**, pp. 115–119
7 Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabási, A.L.: 'The large-scale organization of metabolic networks', *Nature*, 2000, **407**, pp. 651–654
8 Wagner, A., and Fell, D.A.: 'The small world inside large metabolic networks', *Proc. Roy. Soc. London Ser. B-Biol. Sci.*, 2001, **268**, pp. 1803–1810

9 Stuart, J.M., Segal, E., Koller, D., and Kim, S.K.: 'A gene-coexpression network for global discovery of conserved genetic modules', *Science*, 2003, **302**, pp. 249–255
10 Valencia, A., and Pazos, F.: 'Computational methods for the prediction of protein interactions', *Curr. Opin. Struct. Biol.*, 2002, **12**, pp. 368–373
11 Tong, A.H., Lesage, G., Bader, G.D., Ding, H., and Xu, H. *et al.*: 'Global mapping of the yeast genetic interaction network', *Science*, 2004, **303**, pp. 808–813
12 Paulsson, B.: 'Systems biology' (Cambridge University Press, 2006)
13 Alon, U.: 'An introduction to systems biology: design principles of biological circuits' (Chapman and Hall, 2006)
14 Sako, Y.: 'Imaging single molecules in living cells for systems biology', *Mol. Syst. Biol.*, 2006, **2**, pp. 56–62
15 Tyson, J.J., Chen, K.C., and Novak, B.: 'Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell', *Curr. Opin. Cell. Biol.*, 2003, **15**, pp. 221–231
16 Papin, J.A., Hunter, T., Palsson, B.O., and Subramaniam, S.: 'Reconstruction of cellular signalling networks and analysis of their properties', *Nat. Rev. Mol. Cell. Biol.*, 2005, **6**, pp. 99–111
17 Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., and Basrai, M.A. *et al.*: 'Characterization of the yeast transcriptome', *Cell*, 1997, **88**, pp. 243–251
18 Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., and Gallo, M.V. *et al.*: 'Expression monitoring by hybridization to high-density oligonucleotide arrays', *Nat. Biotechnol.*, 1996, **14**, pp. 1675–1680
19 DeRisi, J.L., Iyer, V.R., and Brown, P.O.: 'Exploring the metabolic and genetic control of gene expression on a genomic scale', *Science*, 1997, **278**, pp. 680–686
20 Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., and Conway, A. *et al.*: 'A genome-wide transcriptional analysis of the mitotic cell cycle', *Mol. Cell*, 1998, **2**, pp. 65–73
21 Wen, X., Fuhrman, S., Michaels, G.S., Carr, D.B., and Smith, S. *et al.*: 'Large-scale temporal gene expression mapping of central nervous system development', *Proc. Natl. Acad. Sci. USA*, 1998, **95**, pp. 334–339
22 Hasty, J., McMillen, D., Isaacs, F., and Collins, J.J.: 'Computational studies of gene regulatory networks: in numero molecular biology', *Nat. Rev. Genet.*, 2001, **2**, pp. 268–279
23 Fraga, M.F., Uriol, E., Borja Diego, L., Berdasco, M., and Esteller, M. *et al.*: 'High-performance capillary electrophoretic method for the quantification of 5-methyl 2'-deoxycytidine in genomic DNA: application to plant, animal and human cancer tissues', *Electrophoresis*, 2002, **23**, pp. 1677–1681
24 Galas, D.J., and Schmitz, A.: 'DNAse footprinting: a simple method for the detection of protein-DNA binding specificity', *Nucleic Acids Res.*, 1978, **5**, pp. 3157–3170
25 Kuo, M.H., and Allis, C.D.: 'In vivo cross-linking and immunoprecipitation for studying dynamic Protein: DNA associations in a chromatin environment', *Methods*, 1999, **19**, pp. 425–433
26 Horak, C.E., Luscombe, N.M., Qian, J., Bertone, P., and Piccirrillo, S. *et al.*: 'Complex transcriptional circuitry at the G1/S transition in Saccharomyces cerevisiae', *Genes Dev.*, 2002, **16**, pp. 3017–3033
27 Tavazoie, S., and Church, G.M.: 'Quantitative whole-genome analysis of DNA-protein interactions by in vivo methylase protection in E. coli', *Nat. Biotechnol.*, 1998, **16**, pp. 566–571
28 van Steensel, B., and Henikoff, S.: 'Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase', *Nat. Biotechnol.*, 2000, **18**, pp. 424–428
29 Sun, L.V., Chen, L., Greil, F., Negre, N., and Li, T.R. *et al.*: 'Protein-DNA interaction mapping using genomic tiling path microarrays in Drosophila', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, pp. 9428–9433
30 Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., and Snyder, M. *et al.*: 'Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF', *Nature*, 2001, **409**, pp. 533–538
31 Wingender, E., Dietze, P., Karas, H., and Knuppel, R.: 'TRANSFAC: a database on transcription factors and their DNA binding sites', *Nucleic Acids Res.*, 1996, **24**, pp. 238–241
32 Huerta, A.M., Salgado, H., Thieffry, D., and Collado-Vides, J.: 'RegulonDB: a database on transcriptional regulation in Escherichia coli', *Nucleic Acids Res.*, 1998, **26**, pp. 55–59
33 Kanehisa, M., and Goto, S.: 'KEGG: Kyoto encyclopedia of genes and genomes', *Nucleic Acids Res.*, 2000, **28**, pp. 27–30
34 Monti, M., Orru, S., Pagnozzi, D., and Pucci, P.: 'Interaction proteomics', *Biosci. Rep.*, 2005, **25**, pp. 45–56
35 Ashman, K., Moran, M.F., Sicheri, F., Pawson, T., and Tyers, M.: 'Cell signalling – the proteomics of it all', *Sci. STKE*, 2001, pp. PE33
36 Kumar, A., and Snyder, M.: 'Protein complexes take the bait', *Nature*, 2002, **415**, pp. 123–124
37 Puig, O., Caspary, F., Rigaut, G., Rutz, B., and Bouveret, E. *et al.*: 'The tandem affinity purification (TAP) method: A general procedure of protein complex purification', *Methods*, 2000, **24**, pp. 218–229
38 Foster, L.J., De Hoog, C.L., and Mann, M.: 'Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, pp. 5813–5818
39 Zhu, H., and Snyder, M.: 'Protein chip technology', *Curr. Opin. Chem. Biol.*, 2003, **7**, pp. 55–63
40 Aebersold, R., and Mann, M.: 'Mass spectrometry-based proteomics', *Nature*, 2003, **422**, pp. 198–207
41 Blagoev, B., Kratchmarova, I., Ong, S.E., Nielsen, M., and Foster, L.J. *et al.*: 'A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling', *Nat. Biotechnol.*, 2003, **21**, pp. 315–318
42 Schulze, W.X., and Mann, M.: 'A novel proteomic screen for peptide-protein interactions', *J. Biol. Chem.*, 2004, **279**, pp. 10756–10764
43 Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., and Moore, L. *et al.*: 'Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry', *Nature*, 2002, **415**, pp. 180–183
44 Gavin, A.C., Bosche, M., Krause, R., Grandi, P., and Marzioch, M. *et al.*: 'Functional organization of the yeast proteome by systematic analysis of protein complexes', *Nature*, 2002, **415**, pp. 141–147
45 Giot, L., Bader, J.S., Brouwer, C., Chaudhuri, A., and Kuang, B. *et al.*: 'A protein interaction map of Drosophila melanogaster', *Science*, 2003, **302**, pp. 1727–1736
46 Li, S., Armstrong, C.M., Bertin, N., Ge, H., and Milstein, S. *et al.*: 'A map of the interactome network of the metazoan C. elegans', *Science*, 2004, **303**, pp. 540–543
47 Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., and Dricot, A. *et al.*: 'Towards a proteome-scale map of the human protein-protein interaction network', *Nature*, 2005, **437**, pp. 1173–1178
48 Zhu, H., Bilgin, M., Bangham, R., Hall, D., and Casamayor, A. *et al.*: 'Global analysis of protein activities using proteome chips', *Science*, 2001, **293**, pp. 2101–2105
49 Sidhu, S.S., Bader, G.D., and Boone, C.: 'Functional genomics of intracellular peptide recognition domains with combinatorial biology methods', *Curr. Opin. Chem. Biol.*, 2003, **7**, pp. 97–102
50 Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., and Kim, S.M. *et al.*: 'DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions', *Nucleic Acids Res.*, 2002, **30**, pp. 303–305
51 Bader, G.D., Donaldson, I., Wolting, C., Ouellette, B.F., and Pawson, T. *et al.*: 'BIND – the biomolecular interaction network database', *Nucleic Acids Res.*, 2001, **29**, pp. 242–245
52 Mewes, H.W., Amid, C., Arnold, R., Frishman, D., and Guldener, U. *et al.*: 'MIPS: analysis and annotation of proteins from whole genomes', *Nucleic Acids Res.*, 2004, **32**, D41–44
53 Peri, S., Navarro, J.D., Kristiansen, T.Z., Amanchy, R., and Surendranath, V. *et al.*: 'Human protein reference database as a discovery resource for proteomics', *Nucleic Acids Res.*, 2004, **32**, D497–501
54 Garrels, J.I.: 'YPD-A database for the proteins of Saccharomyces cerevisiae', *Nucleic Acids Res.*, 1996, **24**, pp. 46–49
55 Bussow, K., Cahill, D., Nietfeld, W., Bancroft, D., and Scherzinger, E. *et al.*: 'A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library', *Nucleic Acids Res.*, 1998, **26**, pp. 5007–5008
56 Bussow, K., Nordhoff, E., Lubbert, C., Lehrach, H., and Walter, G.: 'A human cDNA library for high-throughput protein expression screening', *Genomics*, 2000, **65**, pp. 1–8
57 Liu, L.X., Spoerke, J.M., Mulligan, E.L., Chen, J., and Reardon, B., *et al.*: 'High-throughput isolation of Caenorhabditis elegans deletion mutants', *Genome Res.*, 1999, **9**, pp. 859–867
58 Raamsdonk, L.M., Teusink, B., Broadhurst, D., Zhang, N., and Hayes, A. *et al.*: 'A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations', *Nat. Biotechnol.*, 2001, **19**, pp. 45–50
59 Maharjan, R.P., and Ferenci, T.: 'Global metabolite analysis: the influence of extraction methodology on metabolome profiles of Escherichia coli', *Anal. Biochem.*, 2003, **313**, pp. 145–154
60 Wiechert, W., Mollney, M., Isermann, N., Wurzel, W., and de Graaf, A.A.: 'Bidirectional reaction steps in metabolic networks. III. Explicit solution and analysis of isotopomer labeling systems', *Biotechnol. Bioeng.*, 1999, **66**, pp. 69–85
61 Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., and Tomita, M. *et al.*: 'Quantitative metabolome analysis using capillary electrophoresis mass spectrometry', *J. Proteome Res.*, 2003, **2**, pp. 488–494
62 Soga, T., Ueno, Y., Naraoka, H., Matsuda, K., and Tomita, M. *et al.*: 'Pressure-assisted capillary electrophoresis electrospray ionization

mass spectrometry for analysis of multivalent anions', *Anal. Chem.*, 2002, **74**, pp. 6224–6229
63 Christensen, B., and Nielsen, J.: 'Metabolic network analysis of Penicillium chrysogenum using (13)C-labeled glucose', *Biotechnol. Bioeng.*, 2000, **68**, pp. 652–659
64 Wittmann, C., and Heinzle, E.: 'Application of MALDI-TOF MS to lysine-producing Corynebacterium glutamicum: a novel approach for metabolic flux analysis', *Eur. J. Biochem.*, 2001, **268**, pp. 2441–2455
65 Kelleher, J.K.: 'Flux estimation using isotopic tracers: common ground for metabolic physiology and metabolic engineering', *Metab. Eng.*, 2001, **3**, pp. 100–110
66 Lindenthal, B., Aldaghlas, T.A., Holleran, A.L., Sudhop, T., and Berthold, H.K. *et al.*: 'Isotopomer spectral analysis of intermediates of cholesterol synthesis in human subjects and hepatic cells', *Am. J. Physiol. Endocrinol. Metab.*, 2002, **282**, pp. E1222–E1230
67 Christiansen, T., Christensen, B., and Nielsen, J.: 'Metabolic network analysis of Bacillus clausii on minimal and semirich medium using (13)C-labeled glucose', *Metab. Eng.*, 2002, **4**, pp. 159–169
68 Fischer, E., and Sauer, U.: 'Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS', *Eur. J. Biochem.*, 2003, **270**, pp. 880–891
69 Rapoport, S.I.: 'In vivo approaches and rationale for quantifying kinetics and imaging brain lipid metabolic pathways', *Prostaglandins Other Lipid Mediat.*, 2005, **77**, pp. 185–196
70 Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., and Paley, S.M. *et al.*: 'The EcoCyc and MetaCyc databases', *Nucleic Acids Res.*, 2000, **28**, pp. 56–59
71 Dougherty, E.R., Barrera, J., Brun, M., Kim, S., and Cesar, R.M. *et al.*: 'Inference from clustering with application to gene-expression microarrays', *J. Comput. Biol.*, 2002, **9**, pp. 105–126
72 Qian, J., Dolled-Filhart, M., Lin, J., Yu, H., and Gerstein, M.: 'Beyond synexpression relationships: local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions', *J. Mol. Biol.*, 2001, **314**, pp. 1053–1066
73 Marcotte, E.M., Xenarios, I., and Eisenberg, D.: 'Mining literature for protein-protein interactions', *Bioinformatics*, 2001, **17**, pp. 359–363
74 Stapley, B.J., and Benoit, G.: 'Biobibliometrics: information retrieval and visualization from co-occurrences of gene names in Medline abstracts'. Pac. Symp. on Biocomputing, Honolulu, Hawaii, 2000, pp. 529–540
75 Drawid, A., and Gerstein, M.: 'A Bayesian system integrating expression data with sequence patterns for localizing proteins: comprehensive application to the yeast genome', *J. Mol. Biol.*, 2000, **301**, pp. 1059–1075
76 Jansen, R., Yu, H., Greenbaum, D., Kluger, Y., and Krogan, N.J. *et al.*: 'A Bayesian networks approach for predicting protein-protein interactions from genomic data', *Science*, 2003, **302**, pp. 449–453
77 Toronen, P., Kolehmainen, M., Wong, G., and Castren, E.: 'Analysis of gene expression data using self-organizing maps', *FEBS Lett.*, 1999, **451**, pp. 142–146
78 Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M.: 'Systematic determination of genetic network architecture', *Nat. Genet.*, **22**, pp. 281–285
79 Gupta, A., Maranas, C.D., and Albert, R.: 'Elucidation of directionality for co-expressed genes: predicting intra-operon termination sites', *Bioinformatics*, 2006, **22**, pp. 209–214
80 Horvath, S., Zhang, B., Carlson, M., Lu, K.V., and Zhu, S. *et al.*: 'Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target', *Proc. Natl. Acad. Sci. USA*, 2006, **103**, pp. 17402–17407
81 Ge, H., Liu, Z., Church, G.M., and Vidal, M.: 'Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae', *Nat. Genet.*, 2001, **29**, pp. 482–486
82 Alon, U., Barkai, N., Notterman, D.A., Gish, K., and Ybarra, S., *et al.*: 'Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays', *Proc. Natl. Acad. Sci. USA*, 1999, **96**, pp. 6745–6750
83 Gargalovic, P.S., Imura, M., Zhang, B., Gharavi, N.M., and Clark, M.J. *et al.*: 'Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids', *Proc. Natl. Acad. Sci. USA*, 2006, **103**, pp. 12741–12746
84 Friedman, C., Kra, P., Yu, H., Krauthammer, M., and Rzhetsky, A.: 'GENIES: a natural-language processing system for the extraction of molecular pathways from journal articles', *Bioinformatics*, 2001, **17**, (Suppl. 1), pp. S74–82
85 Snel, B., Lehmann, G., Bork, P., and Huynen, M.A.: 'STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene', *Nucleic Acids Res.*, 2000, **28**, pp. 3442–3424
86 von Mering, C.., Jensen, L.J., Kuhn, M., Chaffron, S., and Doerks, T., *et al.*: 'STRING 7-recent developments in the integration and prediction of protein interactions', *Nucleic Acids Res.*, 2007, **35**, pp. D358–D362
87 Friedman, N., Linial, M., Nachman, I., and Pe'er, D.: 'Using Bayesian networks to analyze expression data', *J. Comput. Biol.*, 2000, **7**, pp. 601–620
88 Yu, J., Smith, V.A., Wang, P.P., Hartemink, A.J., and Jarvis, E.D.: 'Advances to Bayesian network inference for generating causal networks from observational biological data', *Bioinformatics*, 2004, **20**, pp. 3594–3603
89 Chen, T., He, H.L., and Church, G.M.: 'Modeling gene expression with differential equations'. Pac Symp. on Biocomputing, Maui, Hawaii, 1999, pp. 29–40
90 Gupta, A., Varner, J.D., and Maranas, C.D.: 'Large-sale inference of the transcriptional regulation of Bacillus subtilis', *Comput. Chemi Eng.*, 2005, **29**, pp. 565–576
91 Shmulevich, I., Dougherty, E.R., Kim, S., and Zhang, W.: 'Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks', *Bioinformatics*, 2002, **18**, pp. 261–274
92 Smolen, P., Baxter, D.A., and Byrne, J.H.: 'Mathematical modeling of gene networks', *Neuron*, 2000, **26**, pp. 567–580
93 Dougherty, E.R., Kim, S., and Chen, Y.: 'Coefficient of determination in nonlinear signal processing', *Signal Process.*, 2000, **80**, pp. 2219–2235
94 Liang, S., Fuhrman, S., and Somogyi, R.: 'Reveal, a general reverse engineering algorithm for inference of genetic network architectures'. Pac Symp. on Biocomputing, Maui, Hawaii, 1998, pp. 18–29
95 Famili, I., Forster, J., Nielsen, J., and Palsson, B.O.: 'Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, pp. 13134–13139
96 Famili, I., and Palsson, B.O.: 'The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools', *Biophys. J.*, 2003, **85**, pp. 16–26
97 Reed, J.L., Vo, T.D., Schilling, C.H., and Palsson, B.O.: 'An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR)', *Genome Biol.*, 2003, **4**, pp. R54
98 Irvine, D.H., and Savageau, M.A.: 'Efficient solution of nonlinear ODE's expressed in S-system canonical form', *SIAM J. Numer. Anal.*, 1990, **27**, pp. 704–735
99 Burgard, A.P., Pharkya, P., and Maranas, C.D.: 'Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization', *Biotechnol. Bioeng.*, 2003, **84**, pp. 647–657
100 Herrgard, M.J., Fong, S.S., and Palsson, B.O.: 'Identification of genome-scale metabolic network models using experimentally measured flux profiles', *PLoS Comput. Biol.*, 2006, **2**, p. e72
101 Arkin, A., Shen, P., and Ross, J.: 'A test case of correlation metric construction of a reaction pathway from measurements', *Science*, 1997, **277**, pp. 1275–1279
102 Dougherty, E.R., and Shmulevich, I.: 'Mappings between probabilistic Boolean networks', *Signal Process.*, 2003, **83**, pp. 799–809
103 Zhao, W., Serpedin, E., and Dougherty, E.R.: 'Inferring gene regulatory networks from time series data using the minimum description length principle', *Bioinformatics*, 2006, **22**, (17), pp. 2129–2135
104 Li, S., Assmann, S.M., and Albert, R.: 'Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling', *PLoS Biol.*, 2006, **4**, pp. e312–329
105 Albert, R., Dasgupta, B., Dondi, R., and Sontag, E.D.: 'Inferring (biological) signal transduction networks via transitive reductions of directed graphs', *Algorithmica*, 2007, in press
106 Wagner, A.: 'Reconstructing pathways in large genetic networks from genetic perturbations', *J. Comput. Biol.*, 2004, **11**, pp. 53–60
107 Albert, R., and Barabási, A.-L.: 'Statistical mechanics of complex networks', *Revi. Mod. Phys.*, 2002, **74**, pp. 47–97
108 Guelzim, N., Bottani, S., Bourgine, P., and Kepes, F.: 'Topological and causal structure of the yeast transcriptional regulatory network', *Nature Genet.*, 2002, **31**, pp. 60–63
109 Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L.: 'The large-scale organization of metabolic networks', *Nature*, 2000, **407**, pp. 651–654
110 Watts, D., and Strogatz, S.H.: 'Collective dynamics of 'small-world' networks', *Nature*, 1998, **393**, pp. 440–442

111 Ravasz, R., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabasi, A.L.: 'Hierarchical organization of modularity in metabolic networks', *Science*, 2002, **297**, pp. 1551–1555
112 Yook, S.H., Oltvai, Z.N., and Barabási, A.L.: 'Functional and topological characterization of protein interaction networks', *Proteomics*, 2004, **4**, pp. 928–942
113 Bollobás, B.: 'Graph theory: an introductory course' (Springer-Verlag, 1979)
114 Dijkstra, E.W.: 'A note on two problems in connection with graphs', *Numer. Math*, 1959, **1**, pp. 269–271
115 Latora, V., and Marchiori, M.: 'Efficient behavior of small-world networks', *Phys. Rev. Lett.*, 2001, **87**, p. 198701–198704
116 Latora, V., and Marchiori, M.: 'Economic small-world behavior in weighted networks', *Eur. Phys. J.*, 2003, **B32**, pp. 249–263
117 Ma'ayan, A., Blitzer, R.D., and Iyengar, R.: 'Toward predictive models of mammalian cells', *Ann. Rev. Giophys. Biomol. Struct.*, 2004, **34**, pp. 319–349
118 Ma'ayan, A., Jenkins, S.L., Neves, S., Hasseldine, A., and Grace, E. *et al*.: 'Formation of regulatory patterns during signal propagation in a Mammalian cellular network', *Science*, 2005, **309**, pp. 1078–1083
119 Anthonisse, J.M.: 'The rush in a directed graph' (Stichting Mathematisch Centrum, Amsterdam, 1971)
120 Freeman, C.L.: 'A set of measures of centrality based on betweenness', *Sociometry*, 1977, **40**, pp. 35–41
121 Girvan, M., and Newman, M.E.J.: 'Community structure in social and biological networks', *Proc. Natl. Acad. Sci. USA*, 2002, **99**, pp. 7821–7826
122 Goh, K.I., Kahng, B., and Kim, D.: 'Universal behavior of load distribution in scale-free networks', *Phys. Rev. Lett.*, 2001, **87**, pp. 278701–278706
123 Holme, P., Huss, M., and Jeong, H.: 'Subnetwork hierarchies of biochemical pathways', *Bioinformatics*, 2003, **19**, pp. 532–538
124 Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U.: 'Network motifs in the transcriptional regulation network of Escherichia coli', *Nat. Genet.*, 2002, **31**, pp. 64–68
125 Luscombe, N.M., Babu, M.M., Yu, H.Y., Snyder, M., and Teichmann, S.A. *et al*.: 'Genomic analysis of regulatory network dynamics reveals large topological changes', *Nature*, 2004, **431**, pp. 308–312
126 McCraith, S., Holtzman, T., Moss, B., and Fields, S.: 'Genome-wide analysis of vaccinia virus protein-protein interactions', *Proc. Natl. Acad. Sci. USA*, 2000, **97**, pp. 4879–4884
127 Rain, J.C., Selig, L., De Reuse, Battaglia, V., and Reverdy, C. *et al*.: 'The protein-protein interaction map of Helicobacter pylori', *Nature*, 2001, **409**, pp. 211–215
128 Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., and Judson, R.S. *et al*.: 'A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae', *Nature*, 2000, **403**, pp. 623–627
129 Balázsi, G., Barabási, A.L., and Oltvai, Z.N.: 'Topological units of environmental signal processing in the transcriptional regulatory network of Escherichia coli', *Proc. Natl. Acad. Sci. USA*, 2005, **102**, pp. 7841–7846
130 Jeong, H., Mason, S.P., Barabási, A.L., and Oltvai, Z.N.: 'Lethality and centrality in protein networks', *Nature*, 2001, **411**, pp. 41–42
131 Maslov, S., and Sneppen, K.: 'Specificity and stability in topology of protein networks', *Science*, 2002, **296**, pp. 910–913
132 Tanaka, R.: 'Scale-rich metabolic networks', *Phys. Rev. Lett.*, 2005, **94**, p. 168101
133 Arita, M.: 'The metabolic world of Escherichia coli is not small', *Proc. Natl. Acad. Sci. USA*, 2004, **101**, pp. 1543–1547
134 Ma, H.W., and Zeng, A.P.: 'The connectivity structure, giant strong component and centrality of metabolic networks', *Bioinformatics*, 2003, **19**, pp. 1423–1430
135 Giaever, G., Chu, A.M., Ni, L., Connelly, C., and Riles, L. *et al*.: 'Functional profiling of the Saccharomyces cerevisiae genome', *Nature*, 2002, **418**, pp. 387–391
136 Said, M.R., Begley, T.J., Oppenheim, A.V., Lauffenburger, D.A., and Samson, L.D.: 'Global network analysis of phenotypic effects: protein networks and toxicity modulation in Saccharomyces cerevisiae', *Proc. Natl. Acad. Sci. USA*, 2004, **101**, pp. 18006–18011
137 Vogelstein, B., Lane, D., and Levine, A.J.: 'Surfing the p53 network', *Nature*, 2000, **408**, pp. 307–310
138 Mahadevan, R., and Palsson, B.O.: 'Properties of metabolic networks: structure versus function', *Biophys. J.*, 2005, **88**, L07–09
139 Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W.: 'From molecular to modular cell biology', *Nature*, 1999, **402**, pp. C47–C52
140 Rives, A.W., and Galitski, T.: 'Modular organization of cellular networks', *Proc. Natl. Acad. Sci USA*, 2003, **100**, pp. 1128–1133
141 Spirin, V., and Mirny, L.A.: 'Protein complexes and functional modules in molecular networks', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, pp. 12123–12128
142 Guimera, R., and Nunes Amaral, L.A.: 'Functional cartography of complex metabolic networks', *Nature*, 2005, **433**, pp. 895–900
143 Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., and Berriz, G.F. *et al*.: 'Evidence for dynamically organized modularity in the yeast protein-protein interaction network', *Nature*, 2004, **430**, pp. 88–93
144 Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabási, A.L.: 'Hierarchical organization of modularity in metabolic networks', *Science*, 2002, **297**, pp. 1551–1555
145 Wuchty, S., Oltvai, Z.N., and Barabasi, A.L.: 'Evolutionary conservation of motif constituents in the yeast protein interaction network', *Nat. Genet.*, 2003, **35**, pp. 176–179
146 Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., and Milo, R. *et al*.: 'Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction', *Proc. Natl. Acad. Sci. USA*, 2004, **101**, pp. 5934–5939
147 Zhang, L.V., King, O.D., Wong, S.L., Goldberg, D.S., and Tong, A.H. *et al*.: 'Motifs, themes and thematic maps of an integrated Saccharomyces cerevisiae interaction network', *J. Biol.*, 2005, **4**, pp. 6–19
148 Mangan, S., Zaslaver, A., and Alon, U.: 'The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks', *J. Mol. Biol.*, 2003, **334**, pp. 197–204
149 Kalir, S., Mangan, S., and Alon, U.: 'A coherent feed-forward loop with a SUM input function prolongs flagella expression in Escherichia coli', *Mol. Syst. Biol.*, 2005, **1**, pp. 2005–2006
150 Papin, J.A., and Palsson, B.O.: 'Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk', *J. Theor. Biol.*, 2004, **227**, pp. 283–297
151 Erdos, P., and Rényi, A.: 'on the evolution of random graphs', *Bull. Inst. Int. Stat.*, 1961, **38**, pp. 343–347
152 Newman, M.E.J., and Watts, D.: 'Scaling and percolation in the small-world network model', *Phys. Rev. E.*, 1999, **60**, pp. 7332–7342
153 Cohen, R., Havlin, S., and ben-Avraham, D.: 'Structural properties of scale free networks' in Borhholdt, S. and Schuster, H.G. (Eds.): 'Handbook of graphs and networks' (Wiley-VCH, 2002), pp. 85–110
154 Newman, M.E.J.: 'Random graphs as models of networks' in Borhholdt, S. and Schuster, H.G. (Eds.): 'Handbook of graphs and networks', (Wiley-VCH, 2002), pp. 35–65
155 Milo, R., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U.: 'Network motifs: simple building blocks of complex networks', *Science*, 2002, **298**, pp. 824–827
156 Shen-Orr, S., Milo, R., Mangan, S., and Alon, U.: 'Network motifs in the transcriptional regulation network of Escherichia coli', *Nat. Genet.*, 2002, **31**, pp. 64–68
157 Barabási, A.-L., and Albert, R.: 'Emergence of scaling in random networks', *Science*, 1999, **286**, pp. 509–512
158 Bollobás, B., and Riordan, O.M.: 'Mathematical results on scale-free random graphs' in Borhholdt, S. and Schuster, H.G. (Eds.): 'Handbook of graphs and networks', (Wiley-VCH, 2002), pp. 1–32
159 Krapivsky, P.L., Redner, S., and Leyvraz, F.: 'Connectivity of growing random networks', *Phys. Rev. Lett.*, 2000, **85**, pp. 4629–4632
160 Dorogovtsev, S.N., Mendes, J.F., and Samukhin, A.N.: 'Structure of growing networks with preferential linking', *Phys. Rev. Lett.*, 2000, **85**, pp. 4633–4636
161 Dorogovtsev, S.N., and Mendes, J.F.F.: 'Accelerated growth of networks' in Borhholdt, S. and Schuster, H.G.'Handbook of graphs and networks', (Wiley-VCH, 2002), pp. 320–343
162 Dorogovtsev, S.N., and Mendes, J.F.F.: 'Evolution of networks', *Adv. Phys.*, 2002, **51**, pp. 1079–1187
163 Bianconi, G., and Barabási, A.-L.: 'Competition and multiscaling in complex networks', *Europhy. Lett.*, 2001, **54**, pp. 436–442
164 Moore, C., Gourab, G., and Newman, M.E.J.: 'Exact solutions for models of evolving networks with addition and deletion of nodes', *Phys. Rev. E. Stat. Nonlin. Soft Matter Phys.*, 2006, **74**, pp. 036121–036129
165 Newman, M.E.J.: 'The structure and function of complex networks', *SIAM Rev.*, 2003, **45**, pp. 167–256
166 Boccaletti, S., Latora, V., Moreno, M., Chavez, M., and Hwang, D.: 'Complex networks: Structure and dynamics', *Phys. Rep.*, 2006, **424**, pp. 175–308
167 Wagner, A.: 'How the global structure of protein interaction networks evolves', *Proc. Biol. Sci.*, 2003, **270**, pp. 457–66
168 Eisenberg, E., and Levanon, E.Y.: 'Preferential attachment in the protein network evolution', *Phys. Rev. Lett.*, 2003, **91**, pp. 138701–138705
169 Berg, J., Lassig, M., and Wagner, A.: 'Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications', *BMC Evol. Biol.*, 2004, **4**, pp. 51–63
170 Hughes, A.L., and Friedman, R.: 'Gene duplication and the properties of biological networks', *J. Mol. Evol.*, 2005, **61**, pp. 758–764

171 Pastor-Satorras, R., Smith, E., and Sole, R.V.: 'Evolving protein interaction networks through gene duplication', *J. Theor. Biol.*, 2003, **222**, pp. 199–210
172 Kim, J., Krapivsky, P.L., Kahng, B., and Redner, S.: 'Infinite-order percolation and giant fluctuations in a protein interaction network', *Phy. Rev. E*, 2002, **66**, pp. 055101–055105
173 Bersini, H., Lenaerts, T., and Santos, F.C.: 'Growing biological networks: Beyond the gene-duplication model', *J. Theor. Biol.*, 2006, **241**, pp. 488–505
174 Tanay, A., Regev, A., and Shamir, R.: 'Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast', *Proc. Natl. Acad. Sci. USA*, 2005, **102**, pp. 7203–7208
175 Light, S., and Kraulis, P.: 'Network analysis of metabolic enzyme evolution in Escherichia coli', *BMC Bioinformat.*, 2004, **5**, pp. 15–31
176 von Mering, C., Zdobnov, E.M., Tsoka, S., Ciccarelli, F.D., and Pereira-Leal, J.B. *et al.*: 'Genome evolution reveals biochemical networks and functional modules', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, pp. 15428–15488
177 Sharan, R., Suthram, S., Kelley, R.M., Kuhn, T., and McCuine, S. *et al.*: 'Conserved patterns of protein interaction in multiple species', *Proc. Natl. Acad. Sci. USA*, 2005, **102**, pp. 1974–1979
178 Hahn, M.W., Conant, G.C., and Wagner, A.: 'Molecular evolution in large genetic networks: does connectivity equal constraint?', *J. Mol. Evol.*, **58**, pp. 203–211
179 Conant, G.C., and Wagner, A.: 'Convergent evolution of gene circuits', *Nat. Genet.*, 2003, **34**, pp. 264–266
180 Mangan, S., and Alon, U.: 'Structure and function of the feed-forward loop network motif', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, pp. 11980–11985
181 Vergassola, M., Vespignani, A., and Dujon, B.: 'Cooperative evolution in protein complexes of yeast from comparative analyses of its interaction network', *Proteomics*, 2005, **5**, pp. 3116–3119
182 Madan Babu, M., Teichmann, S.A., and Aravind, L.: 'Evolutionary dynamics of prokaryotic transcriptional regulatory networks', *J. Mol. Biol.*, 2006, **358**, pp. 614–633
183 Tringe, S.G., von Mering C., Kobayashi, A., Salamov, A.A., and Chen, K. *et al.*: 'Comparative metagenomics of microbial communities', *Science*, 2005, **308**, pp. 554–557
184 Albert, R., and Othmer, H.G.: 'The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster', *J. Theor. Biol.*, 2003, **223**, pp. 1–18
185 Barkai, N., and Leibler, S.: 'Robustness in simple biochemical networks', *Nature*, 1997, **387**, pp. 913–917
186 von Dassow, G., Meir, M., Munro, E.M., and Odell, G.M.: 'The segment polarity network is a robust developmental module', *Nature*, 2000, **406**, pp. 188–192
187 Andrews, S.S., and Arkin, A.P.: 'Simulating cell biology', *Curr. Biol.*, 2006, **16**, pp. R523–R527
188 Morton-Firth, C.J., and Bray, D.: 'Predicting temporal fluctuations in an intracellular signalling pathway', *J. Theor. Biol.*, 1998, **192**, pp. 117–128
189 Rao, C.V., Wolf, D.M., and Arkin, A.P.: 'Control, exploitation and tolerance of intracellular noise', *Nature*, 2002, **420**, pp. 231–237
190 Shimizu, T.S., Aksenov, S.V., and Bray, D.: 'A spatially extended stochastic model of the bacterial chemotaxis signalling pathway', *J. Mol. Biol.*, 2003, **329**, pp. 291–309
191 Chaves, M., Albert, R., and Sontag, E.D.: 'Robustness and fragility of Boolean models for genetic regulatory networks', *J. Theor. Biol.*, 2005, **235**, pp. 431–449
192 Mendoza, L., and Alvarez-Buylla, E.R.: 'Dynamics of the genetic regulatory network for Arabidopsis thaliana flower morphogenesis', *J. Theor. Biol.*, 1998, **193**, pp. 307–319
193 McAdams, H.H., and Shapiro, L.: 'Circuit simulation of genetic networks', *Science*, 1995, **269**, pp. 650–656
194 Yuh, C.H., Bolouri, H., and Davidson, E.H.: 'Cis-regulatory logic in the endo16 gene: switching from a specification to a differentiation mode of control', *Development*, 2001, **128**, pp. 617–629
195 Chaves, M., Sontag, E.D., and Albert, R.: 'Methods of robustness analysis for Boolean models of gene control networks', *IET Syst. Bial. Syst. Biol.*, 2006, **153**, pp. 154–167
196 Glass, L., and Kauffman, S.A.: 'The logical analysis of continuous, non-linear biochemical control networks', *J. Theor. Biol.*, 1973, **39**, pp. 103–129
197 Albert, I., and Albert, R.: 'Conserved network motifs allow protein-protein interaction prediction', *Bioinformatics*, 2004, **20**, pp. 3346–3352
198 Christensen, C., Gupta, A., Maranas, C.D., and Albert, R.: 'Inference and graph-theoretical analysis of Bacillus Subtilis gene regulatory networks', *Physica A*, 2007, **373**, pp. 796–810
199 Balazsi, G., Barabasi, A.L., and Oltvai, Z.N.: 'Topological units of environmental signal processing in the transcriptional regulatory network of Escherichia coli', *Proc. Natl. Acad. Sci. USA*, 2005, **102**, pp. 7841–7846