Modeling and identification of biological networks

Florence d'Alché–Buc[1] and Vincent Schachter[2]

[1] LaMI UMR CNRS 8042 and Epigenomics Programme, Genopole
523 Place des terrasses
91 000 Evry, France
(e-mail: dalche@lami.univ-evry.fr)
[2] Genoscope, Consortium National de Recherche en Génomique
91 000 Evry, France
(e-mail: vs@genoscope.cns.fr)

Abstract. Modeling, reverse-engineering and analysis of macromolecular networks has spurred increasing interest in the computational biology and the biostatistics communities. Biologists need rigorous and flexible tools to describe, infer and study these complex systems. This survey focuses on some of the latest advances on the corresponding direct and reverse modeling approaches.
Keywords: biological networks, gene networks, metabolic networks, machine learning.

1 Introduction

With the availability of complete genome sequences and high-throughput, post-genomics experimental data, the last 5 years have witnessed a growing interest in the study of networks of macromolecular interactions.

During the last few years, modeling efforts have targeted several distinct types of networks at the molecular level : gene regulatory networks, metabolic networks, signal transduction networks or protein-protein interaction networks, not to mention networks of interactions that are not restricted to a cell (intercellular communications) or take place at an altogether different level of detail (immunological networks, ecological networks). Here, we focus exclusively on molecular processes that take place within a cell, and more specifically on two distinct types of cellular mechanisms : transcriptional regulation and metabolism.

A major challenge consists in identifying with reasonable accuracy those complex macromolecular interactions that take place at different levels from genes to metabolites through proteins. Once identified, a network model can be used to simulate the process it represents, or for a variety of analyses, ranging from statistical properties of its topology to predictions of features of its dynamic behavior, or even prediction of cellular phenotypes.

This review focuses on modeling frameworks for biological networks, and on the existing methods to identify models from data within these frame-

works. Framework choice and design are influenced both by targeted analyses, and by the need for model identification methods that will yield exploitable results given available experimental data and prior knowledge.

2 Models of biological networks

Why design models of biological networks ? A first motivation is to present a synthetic view of the current state of biological knowledge on a given network, and to structure it in a way that brings to sight relevant properties that might remain hidden without the model, or with a less relevant model. A second motivation is to allow predictions of (properties of) the network's dynamical behavior, one key point being that if these predictions can be compared with experimental results, they should allow either confirmation of the model's accuracy or, better yet, correction of the model.

Recent modeling framework proposals abound (see [de Jong, 02] for a detailed review), resulting in significant advances in the biological network modeling field, but also in a conceptual landscape that seems somewhat cluttered and unstructured. This impression is only superficial, however. The landscape can be simplified by regrouping frameworks that have similar underlying mathematical structure. In addition, models are very often goal-oriented, each framework was originally designed with some analytical aim in mind. In the rest of this section, we review families of formalisms classified according to the types of analyses and predictions for which they are best suited.

As we will see in the section 3, however, such a classification is only one-half of the story : available experimental data and model identification methods can also have a strong influence on the choice or design of a framework. The final modeling choice is often the result of a subtle balancing act between the requirements of model identification and the goals of the intended analyses.

2.1 Gene regulatory networks

Transcriptional regulation is the process by which genes regulate the transcription of other genes. A gene A *directly regulates* a gene B if the protein that is encoded by A is a transcription factor for gene B, ie if it binds to DNA on a specific site near the sequence coding for B, called a *regulatory region* of B, and activates or inhibits its rate of transcription. Regulation can be indirect, e.g. A activates B, which activates C, and *cooperative*, i.e. several genes regulate the same target gene in a non purely additive manner.

Several types of experimental data provide information on the transcriptional regulation process, some of which can be produced at high-throughput, while others still result from targeted, context-dependent assays and therefore can be acquired only for small networks. The main high-throughput

technology is DNA chips which measure the concentration of mRNAs (a.k.a the *expression level of genes*) corresponding to all genes (or a large set of genes) in the organism under study, for several time-points or several different *conditions*, i.e. environmental changes, genetic or chemical perturbations of the system [Spellman *et al.*, 98]. These experiments can be seen as providing instantaneous pictures of the state of the regulatory network. Other sources of information include ChIP-chip ([Lee *et al.*, 02]) assays, that detect direct regulatory influences by identifying the binding of a protein to a regulatory region (in other words, a protein-DNA interaction), as well as the identification of sequence segments that are similar to known regulatory sequences, using sequence comparison methods. So far, large-scale expression and protein-DNA datasets have been generated mainly for model organisms, i.e. *Saccharomyces cerevisiae* (common yeast) and to a lesser extent the bacterium *E.coli*, making these the most likely candidates for any large scale reverse modeling effort.

2.1.1 Analyses of network topology : directed graphs The first category of properties of interest in biological networks are those related to their (static) network structure. Such topological analyses are most meaningful when applied to large ('genome-scale') networks, the aim being to identify statistical properties that can be interpreted as 'traces' of underlying biological mechanisms or design principles, related for instance to their dynamics [Shen-Orr *et al.*, 02, Watts and Strogatz, 98] (how the connectivity structure of the biological process reflects its dynamics), to their evolution [Jeong *et al.*, 00, Wagner, 01] (i.e. likely scenarios for the evolution of a network exhibiting the observed property or properties), or to both [Jeong *et al.*, 01, Milo *et al.*, 04].

One should emphasize that those analyses that are motivated by the search for insights into network dynamics focus on network structure mostly because large-scale data on network dynamics is not yet available. They can provide valuable insight insofar as the interpretative leap between static structure and dynamic behavior is performed carefully. Statistical graph properties that have been studied in this context include the distribution of vertex degrees [Jeong *et al.*, 01], the distribution of the clustering coefficient and other notions of density [Newman, 03, Guelzim *et al.*, 02], the distribution of vertex-vertex distances [Ravasz *et al.*, 02], and the distribution of network motifs occurrences [Milo *et al.*, 02].

The framework of choice to study these properties is also the most straightforward one. A gene regulatory network is viewed as a *directed graph* : a pair (V,E) where V is a set of vertices and E a set of directed edges, i.e. pairs (i,j) of vertices, where i is the source vertex and j the target vertex. Vertices of the graph represent genes, edges represent regulatory influences. Note that in some cases, it may be preferable to work with undirected graphs instead,

for instance when only the existence of a correlation between the expression levels of two genes is known, but not the causal direction.

This simple model can be enriched by adding information (labels) on vertices or edges : for instance, '+' or '-' labels on edges may indicate positive or negative regulatory influence, the existence of an edge may be specified as conditional on the cell being in a specific global state, or on the source gene (the regulator) being expressed above a given threshold. These latter types of additional information, however, refer implicitly to notions of state and temporal evolution, and thus lead naturally towards qualitative dynamical models.

Finally, it is worth mentioning that enriched graph representations are also at the core of most existing biological pathways databases [Cary *et al.*, 05]. One reason is their simplicity, another one is that basic or complex queries on biological networks often correspond to classical operations on graphs, e.g. the search for paths between genes obeying given conditions.

2.1.2 Analyses of network dynamics : continuous models, discrete models The dynamics of regulatory processes has been the object of intense recent scrutiny. Whereas understanding the detailed dynamics of a regulatory network requires more experimental information than deciphering its static structure, dynamics is obviously one step closer to biological function.

Models can be used to run simulations of the biological system under study, with various choices of values for parameters corresponding either to unknown system characteristics or to environmental conditions. Comparison of simulated dynamics with experimental measurements can help refine the model or provide insight on qualitative properties of the system's dynamical behavior. The latter can also be addressed directly, by reasoning on or identifying properties of the system's behavior instead of simulating it, with the help of theoretical tools that depend on the choice of formalism. Dynamical properties of interest include the identification of steady states or limit cycles, identification of multistable (e.g. switch-like) behavior , identification of oscillatory behavior, characterization of the role of some parts of the network in terms of signal processing (e.g. amplifiers, derivators, logic gates) , and assessment of robustness environmental changes or genetic perturbation (see [Tyson *et al.*, 03, Wolf and Arkin, 03] for detailed review).

The default modeling option to simulate the dynamics of regulatory processes is to write a system of differential equations that govern the evolution of mRNA and protein concentrations. Typically, a gene regulatory network is modeled as a system of rate equations of the form : $\frac{dx_i}{dt} = f_i(\mathbf{x})$, $1 \leq i \leq n$ where $\mathbf{x}= (x_1, \ldots, x_n)$ is the vector of concentrations (of mRNAs, proteins or small molecules) and $f_i : \Re \rightarrow \Re^n$ a function, not necessarily linear. The level of detail and the complexity of these *kinetic models* can be adjusted, through the choice of the rate functions f_i. Typical tradeoffs include :

- using a more or less simplified set of entities and reactions, e.g. choosing whether to take into account mRNA and protein degradation,
- including delays to account for transcription, translation or diffusion time
- using more or less detailed kinetics, i.e. specific forms of f_i

Systems of differential equations as a modeling framework for biological networks presents two major drawbacks. Each equation in the model requires the knowledge of one or several parameter values (thermodynamic constants, rate constants), which is out of the present reach of high-throughput data production techniques. It is thus difficult to instantiate models of large networks directly, and reverse-engineering techniques are limited in how much information they allow to extract from limited datasets. Moreover, deriving meaningful dynamical properties of large differential equations system is a challenge : the f_i being nonlinear, analytical solutions are not known in the general case. So far these systems have been mainly used for numerical simulations within given parameter ranges (realistic or not), possibly complemented by bifurcation analysis, rather than submitted to analytical approaches [de Jong, 02].

These limitations have motivated two main tracks of investigation on alternative modeling frameworks for biological networks : simplified kinetic models on one hand, and discrete[1] models on the other hand.

Simplified continuous frameworks include piecewise-linear differential equations, a special case of rate equations where the response of a gene to regulatory stimuli (the function f_i) is approximated by a step function [de Jong, 02]. Linearity facilitates the analytical treatment of some dynamical properties, such as steady states. Systems of piecewise-linear differential equations can also be analyzed qualitatively by discretizing and recasting them within the framework of *qualitative differential equations*, where variables and their derivatives take qualitative (discrete) values and functions f_i are abstracted into sets of qualitative constraints.

Several discrete modeling frameworks have been proposed, each with a specific tradeoff between the level of detail of its chosen observables and the type of analyses that it enables : boolean networks (see below), generalized logical networks [Thomas *et al.*, 1995](a generalization of boolean networks that increases biological realism by allowing variables to have more than two values and using asynchronous transitions), petri-nets [Matsuno *et al.*, 2000], process-algebra [Regev *et al.*, 2001], rule-based formalisms [Chabrier *et al.*, 04].

[1] Here, we mean that *time* is discretized, leading to frameworks where the dynamics is governed by state transitions between t and t+1. Discretization of expression levels and/or of rate functions are a different path to simplification of either time-continuous or time-discrete frameworks.

2.2 Metabolic networks

Metabolism is the set of processes by which a cell extracts energy and raw material from its environment, and uses both to produce the components (DNA, proteins, lipids...)necessary for its survival and function, and to interact with its environment. Metabolic networks are thus networks of biochemical reactions : each reaction transforms one or several substrates (metabolites, i.e. small organic molecules) into one or several products (metabolites as well).To occur within a cell at a significant rate, a metabolic reaction needs to be catalyzed by an enzyme (a protein with catalytic activity) specific to that reaction.

Much of the classification introduced above for regulatory networks applies to metabolic networks ; indeed, several formalisms and analytical tools have been used on both. One should not be misled by these similarities, however : metabolic networks and regulatory networks represent very distinct, albeit interrelated, biological mechanisms, and this does translate into mathematical differences.

The framework of choice to capture the connectivity structure of a metabolic network is a directed bipartite graph (rather than a simple directed graph) : vertices correspond respectively to metabolites and reactions, edges represent production or consumption of a metabolite by a reaction. Two types of simpler graphs can be extracted from such a bipartite graph : *enzyme graphs*, where an edge between two reaction vertices denotes the fact that a product of the source reaction is a substrate of the target reaction (and can also denote the causal ordering of reactions in metabolic processes), and *metabolite graphs*, where vertices representing metabolites are linked when a reaction consumes one to produce the other. For all three graph types, active areas of research include the definition of biologically meaningful distances and the design of relevant and computable subgraph similarity measures to allow comparative studies.

Metabolic networks dynamics can be expressed as described above, using systems of rate equations and a given approximation for rate functions. Attempts at analytical reasoning have spurred the development of various simplified frameworks, including *Biochemicals Systems Theory* [Savageau, 1991] where production and consumption rates are expressed using a power-law approximation, and *Metabolic Control Analysis* [Westerhoff *et al.*, 1994], which focuses on a first-order approximation of the dynamical system in the neighborhood of steady-state. It is worth noticing that discrete frameworks have seldom been used to model metabolism : metabolite fluxes and concentrations are the key variables of interest here, in contrast with regulatory networks where an on/off discretization of the state of a gene already provides valuable information on the regulatory logic. Another type of abstract, scalable framework have been successfully applied to metabolic modeling : constraint-based modeling.

Constraint-based modeling is a framework dedicated to the modelling of metabolic processes at steady state : a global state of the metabolic network is defined as a distribution of fluxes within the network reactions. It emerged in the 90s as a simplification of kinetic models (mostly in the Schuster and Palsson groups), and was developed to allow tractable modelling of genome-scale metabolic networks [Price *et al.*, 04]. The steady-state hypothesis positions the framework at a level of detail intermediate between description of static network structure and representation of network dynamics. It is designed to represent incomplete information, yet to allow some prediction of metabolic behavior. The focus, rather than being on fully instantiated descriptions of the system's behavior, is on sets of such descriptions, i.e. sets of flux distributions compatible with a set of constraints representing the current knowledge on the structure of the network, on thermodynamic and kinetic parameters, and on input/output relationships of the network with its environment. The solution set can be refined incrementally as new constraints are added, ensuring some robustness in structural analyses and metabolic behaviour predictions with respect to modifications of the model. As this framework has been applied successfully to a variety structural analyses and predictive tasks on large metabolic networks in bacteria and yeast, yielding interesting biological results, efforts are under way to extend it while preserving simplicity and tractability.

3 Model identification: a machine learning problem

Once a formal framework is defined to describe models of biological networks, the question of how to choose parameters arises. Various works have shown that this identification problem can be expressed in the framework of machine learning. Given a family of mathematical models of gene interactions and a set of observations, learning consists here in optimizing the parameters of the model in such way that it captures the observed behavior of the true system. The ability of the instantiated model to be used in prediction is referred as the generalization property. A model is able to generalize if learning ensures a trade-off between a good fit to the data and simplicity of the model. Solving a learning problem leads to three key questions : the representation problem, the optimization problem and the validation problem. The **representation** problem concerns mostly the choice of the formalism in which data and the model are going to be expressed, and the method to encode them into this formalism. Both symbolic and numerical learning leads to an **optimization** problem whose nature is combinatorial (for symbolic learning) and numeric (for statistical learning). Statistical approaches generally use maximum likelihood criteria penalized by a parsimony constraint. Combinatorial approaches are solved using heuristics to ensure a large exploration of the models spaces. At last **validation** is required to identify how one can trust the inferred model. In this area statistical approaches

benefits from an important background in statistical validation of estimation methods. The validation question is far from being solved in the context of network reconstruction, however.

Most efforts on biological networks identification fall into this framework, albeit most of them do not address each of the above key issues. They can be divided into two categories: on one hand, the static approaches that neglect temporal aspects and focus on the sole reconstruction of the interaction graph, and on the other hand, the dynamic approaches that aim at modeling the underlying dynamic system providing both structure and dynamics parameters.

3.1 Static approaches

Several static approaches have yielded promising results in reconstructing gene networks. In order to focus on the structure identification problem, they ignore the temporal aspects and search for causality chains among the variables at hand. This point of view is based on the implicit assumption that the underlying dynamical process is at equilibrium, and that no circuit exists among studied genes.

Bayesian networks are undoubtedly the most successful approach to gene networks structure reconstruction.They represent the expression levels of genes as random variables, whose joint probability law has to be identified. This model has two major advantages : it takes into account the inherent stochastic character of biological processes and it is able to cope with noisy data. A graphical display of such models can be obtained by considering directed acyclic graphs whose vertices are genes and whose edges are modeled by conditional probabilities distributions. Choosing discrete or continuous variables, parametric or non parametric forms for the cpd's are the main questions in the representation problem.

Learning bayesian networks consists in estimating the joint probability distribution of the variables using available data. The core issue is to find the decomposition of the joint law in the conditional probability distributions (cpd's) among the relevant variables. This decomposition defines the graph structure. Once the structure of a network is given, the task of learning cpd 's is not difficult. Learning the structure, however, is an NP-hard problem that can only be tackled by heuristics. Several pioneering results in this area have been achieved using a constructive strategy. Reconstruction has been shown to be successful on the yeast cell cycle dataset of [Spellman *et al.*, 98].

Extensions of these results were obtained by integration of prior knowledge into the model. For instance, [E.Segal *et al.*, 01] introduced an enriched formalism, probabilistic relational networks (PRN) that allows to deal with object variables instead of simple discrete or continuous variables. Information about promoters and genomic sequence can be thus be introduced. While information propagation in the net is modified and for this reason learning

becomes also more complex, this work opens the door to new formalisms that couple high level descriptions with a probabilistic framework.

Another variations on bayesian networks models is the so-called 'module networks' approach. Module networks introduced in [E.Segal *et al.*, 03] have been proposed as bayesian networks with a special structure, where the variables sharing the same parents are gathered into a so-called module, i.e. a set of genes that appear to be co-regulated in some experimental conditions. Elucidating which are the genes that belong to the same modules and what are the conditions under which these regulations occur can be solved using a Expectation-Maximization(EM) based algorithm that starts from relevant initializations.

Validation can consist in the comparison between the inferred structure and the true structure. The simplest way consists in a comparison between the inferred network with other sources of knowledge. Precision and recall measures, ROC curves have also been proposed to evaluate the power of learning algorithms [Husmeier, 03].

However all these static approaches are not able to discover circuits in a graph interaction. The reason is that without considering time, it is not possible to elucidate feedback interactions that can only be observed an through time.

3.2 Dynamic approaches

Dynamic approaches aim at identifying the dynamics of the system implemented by a biological network while extracting the structure. Only discrete-time models are considered for learning since experimental data come from discrete-point measurements. In the area of genetic networks, the available data take the form of gene expression kinetics measured after some perturbation of the studied organism. While these data are more expensive to generate than static data, a few subsets exist and mainly concerns model organisms such as bacteria or yeast. Modeling dynamics of a network can serve both exploratory and explanatory goals. A long-term goal is of course to exploit these models in simulation and prediction for drug-design and therapeutical targeting. However its should be stresses that this feature has not yet been fully exploited in the existing works.

Dynamic models that have been considered for learning include Boolean networks, artificial recurrent neural networks, dynamic bayesian networks including state-space models. Learning in boolean networks has first been tackled with combinatorial algorithms [Akutsu *et al.*, 1999]and then renewed by using a randomized algorithm. However the best way to reduce complexity of the problem is to reduce the class of boolean functions as proposed in[Gat-Viks *et al.*, 03] with the so-called chain functions. Promising new directions have also been introduced by [Shmulevich *et al.*, 02] with the introduction of Probabilistic Boolean networks and learning algorithms devoted to their reverse engineering.

While boolean networks simplify the description of the system's dynamical behavior, quantitative models have attracted much attention from machine learning community because of the existence of a large set of efficient learning algorithms for numerical data. These models are usually based on the quantization of differential equations. Again the representation issue implies to choose among linear/non linear models and discrete/continuous variables. Keeping the model deterministic allows its implementation as a recurrent artificial network (see for instance [D'haeseleer *et al.*, 00]and [Mjolness *et al.*, 00]) for which learning algorithms such as genetic algorithms and back-propagation through time have been designed .This last feature avoids avoids data-overfitting. This framework can be compared to dynamic Bayesian networks that are inspired from stochastic differential equations [Hoon *et al.*, 03], or can simply be obtained by adding a noise component to the equation [Perrin *et al.*, 03]. These approaches aim at estimating the joint probability of the temporal sequence of network states. The optimization task takes the form of a likelihood maximization problem with a parsimony constraint.

Several dynamic approaches have been applied to different models, first order [Hoon *et al.*, 03], second order models [Perrin *et al.*, 03], and from linear to non linear [Nachman *et al.*, 04] to splines-based models. Validation of dynamical approaches can be performed by measuring the ability of the model to make k-step predictions or to predict the last part of the sequence used for training. However the most difficult point remains the ability of the algorithm to retrieve the structure of the network which can be deduced from the identified parameters.

4 Conclusion and perspectives

We have reviewed modeling formalisms for biological networks and their relationship to down stream analysis and reverse engineering methods. As this field of research matures, it is becoming increaslingly clear that there is no one-size-fits-all solution, but rather a range of frameworks and methods, each with its specific trade-off between abstraction and tractability, the ultimate test being the ability to answer relevant biological questions. Indeed, network models are only starting to become useful tools for biological investigation. Promising research directions include the design of frameworks that allow joint modeling of metabolism and regulation, the refinement of stochastic rule-based frameworks that are meant to capture intrinsic stochasticity in regulatory networks dynamics, the design of dedicated process calculi, and the development of model-checking tools. Another key direction, towards, efficient model inference is the elaboration of formalisms that are able to support high level language of description while managing uncertainty in the data.

References

[Akutsu *et al.*, 1999]T. Akutsu, S. Miyano, and S. Kuhara. Identification of genetic networks from a small number of gene expression patterns under the boolean network model. In *Proc. of PSB*, volume 4, pages 17–28, 1999.

[Cary *et al.*, 05]M. P. Cary, G. D. Bader, and C. Sander. Pathway information for systems biology. *FEBS Lett*, 579(8):1815–20, 05. 0014-5793 Journal Article.

[Chabrier *et al.*, 04]N. Chabrier, M. Chiaverini, V. Danos, F. Fages, and V. Schachter. Modeling and querying biomolecular networks. *Theoretical Computer Science*, 325(1):25–44, 04.

[de Jong, 02]H. de Jong. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol*, 9(1):67–103, 02. 1066-5277 Journal Article Review.

[D'haeseleer *et al.*, 00]P. D'haeseleer, S. Liang, and R. Somogyi. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics*, 16(8):707–726, 00.

[E.Segal *et al.*, 01]E.Segal, T.Taskar andA.Gasch, N.Friedman, and D.Koller. Rich probabilistic models for gene epxression. *Bioinformatics*, 17:S243–252, 01.

[E.Segal *et al.*, 03]E.Segal, D. Pe'er, A. Regev, D. Koller, and N. Friedman. Learning module networks. In *Proc. of UAI*, pages 523–534, 03.

[Gat-Viks *et al.*, 03]I; Gat-Viks, R.Shamir, R.M. Karp, and R. Sharan. Reconstructing chain fucntions in genetic networks. *Bioinformatics*, 19:i108–i117, 03.

[Guelzim *et al.*, 02]N. Guelzim, S. Bottani, P. Bourgine, and F. Kepes. Topological and causal structure of the yeast transcriptional regulatory network. *Nat Genet*, 31(1):60–3, 02.

[Hoon *et al.*, 03]M. De Hoon, S. Imoto, K. Kobayashi, N. Ogasawara, and S. Miyano. Inferring gene regulatory networks form time-ordered gene expression data of bacillus subtilis using differential equations. In *PSB*, pages 17–28, 03.

[Husmeier, 03]Dirk Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic bayesian networks. *Bioinformatics*, 19(17):2271–2282, 03.

[Jeong *et al.*, 00]H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.-L. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 00.

[Jeong *et al.*, 01]H. Jeong, S. P. Mason, A. L. Barabasi, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–2, 01.

[Lee *et al.*, 02]T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young. Transcriptional regulatory networks in saccharomyces cerevisiae. *Science*, 298(5594):799–804, 02.

[Matsuno *et al.*, 2000]H. Matsuno, A. Doi, M. Nagasaki, and S. Miyano. Hybrid petri net representation of gene regulatory network. *Pac Symp Biocomput*, pages 341–52, 2000. Journal Article.

[Milo *et al.*, 02]R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–7, 02.

[Milo *et al.*, 04]R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303(5663):1538–42, 04.

[Mjolness *et al.*, 00]E. Mjolness, T. Mann, R. Castao, and B. Wold. From coexpression to coregulation: An approach to inferring transcriptional regulation among gene classes from large-scale expression data. In *Advances in Neural Information Processing Systems*, volume 12, pages 928–934. 00.

[Nachman *et al.*, 04]I. Nachman, A. Regev, and N. Friedman. Inferring quantitative models of regulatory networks from expression data. *Bioinformatics*, 20:i248–i258, 04.

[Newman, 03]M. E. Newman. Properties of highly clustered networks. *Phys Rev E Stat Nonlin Soft Matter Phys*, 68(2 Pt 2):026121, 03.

[Perrin *et al.*, 03]B.-E. Perrin, L. Ralaivola, A. Mazurie, S. Bottani, and F. d'Alché Buc. Gene networks inference using dynamic bayesian networks. *Bioinformatics*, 19:i38–i49, September 03.

[Price *et al.*, 04]N. D. Price, J. L. Reed, and B. O. Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol*, 2(11):886–97, 04. 1740-1526 Journal Article Review Review, Tutorial.

[Ravasz *et al.*, 02]E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–5, 02.

[Regev *et al.*, 2001]A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the pi-calculus process algebra. *Pac Symp Biocomput*, pages 459–70, 2001. Journal Article.

[Savageau, 1991]M. A. Savageau. Biochemical systems theory: operational differences among variant representations and their significance. *J Theor Biol*, 151(4):509–30, 1991. 0022-5193 Journal Article.

[Shen-Orr *et al.*, 02]S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nat Genet*, 31(1):64–8, 02.

[Shmulevich *et al.*, 02]I. Shmulevich, E. R. Dougherty, S. Kim, and W. Zhang. Probabilistic boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, 18(2):261–274, 02.

[Spellman *et al.*, 98]P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher. Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell*, 9(12):3273–97., 98.

[Thomas *et al.*, 1995]R. Thomas, D. Thieffry, and M. Kaufman. Dynamical behaviour of biological regulatory networks–i. biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull Math Biol*, 57(2):247–76, 1995. 0092-8240 Journal Article.

[Tyson *et al.*, 03]J. J. Tyson, K. C. Chen, and B. Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol*, 15(2):221–31, 03. 0955-0674 Journal Article Review Review, Tutorial.

[Wagner, 01]A. Wagner. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol Biol Evol*, 18(7):1283–92, 01.

[Watts and Strogatz, 98]D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–2, 98.

[Westerhoff *et al.*, 1994]H. V. Westerhoff, J. H. Hofmeyr, and B. N. Kholodenko. Getting to the inside of cells using metabolic control analysis. *Biophys Chem*, 50(3):273–83, 1994. 0301-4622 Journal Article.

[Wolf and Arkin, 03]D. M. Wolf and A. P. Arkin. Motifs, modules and games in bacteria. *Curr Opin Microbiol*, 6(2):125–34, 03.