Article

Topological peculiarities of mammalian networks with different functionalities: transcription, signal transduction and metabolic networks

Björn Goemann, Edgar Wingender, Anatolij P. Potapov

Department of Bioinformatics, University Medical Center Göttingen (UMG), Goldschmidtstrasse 1, D-37077 Göttingen, Germany

E-mail: anatolij.potapov@bioinf.med.uni-goettingen.de

Abstract

We have comparatively investigated three different mammalian networks – on transcription, signal transduction and metabolic processes - with respect to their common and individual topological traits. The networks have been constructed based on genome-wide data collected from human, mouse and rat. None of these three networks exhibits a pure power-law degree distribution and, therefore, could be considered scale-free. Rather, the degree distributions of all three networks were best fitted by mixed models of a power law with an exponential tail. The networks differ from one another in the quantitative parameters of the models. Moreover, the transcription network can also be very well approximated by an exponential law. The connectivity within each network is rather robust, as is seen when removing individual nodes and computing the values of their pairwise disconnectivity index (PDI), which characterizes the topological significance of each node v by the number of direct or indirect connections in the network that critically depend on the presence of v. The results evidence that the networks are not centralized: none of nodes globally controls the integrity of each network. Just a few vertices appeared to strongly affect the coherence of the networks. These nodes are characterized by a broad range of degrees, thereby indicating that the degree alone is not the decisive criteria of a node's importance. The networks reveal distinct architectures: The transcriptional network exhibits a hierarchical modularity, whereas the signaling network is mainly comprised of semi-autonomous modules. The metabolic network seems to be made by a more complex mixture of substructures. Thus, despite being encoded by the same genomes, the networks significantly differ from one another in their general architectural design. Altogether, our results indicate that the subsets of genes and relationships that constitute these networks have co-evolved very differently and through multiple mechanisms.

Keywords mammalian network; transcription; signal transduction; metabolic; topology; modularity.

1 Introduction

In the last decade, intensive studies of the global architecture of various real world networks led to the conclusion that most of them share the small-world property in conjunction with a power-law degree distribution (Barabási and Albert, 1999; Barabási and Oltvai, 2004; Dorogovtsev and Mendes, 2002; Newman, 2003; Albert, 2005; Barabási, 2009). A small-world network is characterized by a small average shortest path length between any two vertices and a large mean clustering coefficient when compared with random networks of the same size (Watts and Strogatz, 1998). A power-law degree distribution is generally perceived as

synonym for its scale-freeness. It implies that most vertices have a very small number of links while few others, so-called "hubs", are highly linked, although other degree distributions (e.g. exponential decays) exhibit this property as well (Tacutu et al., 2011; Zhang, 2011). It is an important characteristic of such a topology that it assures an amazing robustness against random failures and is sensitive only to targeted attacks on the hub nodes (Barabási and Albert, 1999; Albert et al., 2000; Jeong et al., 2001; Barabási and Oltvai, 2004). They are thought to connect various modules inside a scale-free network, i.e. sets of highly interlinked vertices, which themselves have been reported to be organized in a hierarchical way (Barabási and Oltvai, 2004; Newman, 2002; Girvan and Newman, 2002). Most notable, however, is that an evolutionary model has been reported that explains well how networks with a scale-free degree distribution have been generated, i.e. through a preferential attachment of newly emerging vertices to hubs (Barabási and Albert, 1999; Barabási and Oltvai, 2004; Albert, 2005).

So far, such features have been reported for different types of biological networks in various organisms: they range from metabolic networks in bacteria, archaea and eukaryotes (Ravasz et al., 2002; Wagner and Fell, 2001; Ma and Zeng, 2003; Wagner, 2001) over protein-protein interaction networks in yeast and fly (Jeong et al., 2001; Maslov and Sneppen, 2002; Yook et al., 2004; Giot et al., 2003) to some signaling (Papin and Palsson, 2004; Papin et al., 2005) and eukaryotic gene expression networks (Featherstone and Broadie, 2002; Agrawal, 2002; Bhan et al., 2002; Carter et al., 2004; Luscombe et al., 2004). For regulatory networks in higher eukaryotes, only scarce information is available so far about their large-scale characteristics (Ma'ayan et al., 2005; Potapov et al., 2005; Goemann et al., 2009a, Goemann et al., 2009b; Bhardwaj et al., 2010; Ibrahim et al., 2011; Martínez-Antonio, 2011). The current knowledge about the global organization of biological networks is therefore almost completely based on the analysis of prokaryotes and unicellular eukaryotes. It remains to be seen whether the architecture of the corresponding networks in multicellular systems considerably deviates from that in unicellular ones. Since the processes in higher eukaryotes require a much higher regulatory overhead to coordinate the differential gene expression programs in various cell types and tissues one may expect several peculiarities in the corresponding networks.

This work extends our previous studies on selected mammalian networks (Potapov et al., 2005; Goemann et al., 2009a, Goemann et al., 2009b) by describing and comparing the general topologies of mammalian regulatory systems. Each of the inspected networks displays a specific functional aspect of a mammalian cell: (*i*) the transcription network, representing the relationships between transcription factor genes, (*ii*) the signaling network, comprising all known signal transduction molecules, and (*iii*) the metabolic network composed of genes encoding metabolic enzymes. Our previous studies have already indicated a peculiarity of the mammalian transcription network compared with the signaling and the metabolic network in being void of 3-node feed-back loops (Goemann et al., 2009a, Goemann et al., 2009b). The focus here is on their degree distributions, modular organization, and robustness against random and targeted perturbations. We show that despite being encoded by the same genomes, these networks significantly differ from one another in their general architectural design.

2 Methods

2.1 Estimation of the degree distribution

Estimation of the parameters of the power-law, exponential law and power-law with exponential tail distributions have been calculated using the maximum likelihood method. In the case of the power-law, $P(k) \approx k^{-\gamma} / \zeta(\gamma)$ where $\zeta(\gamma)$ is the Riemann zeta function, i.e. $\zeta(\gamma) = \sum_{k=1}^{\infty} k^{-\gamma}$. The likelihood function is given by $L(\gamma \mid k) = \prod_{k=1}^{N} k^{-\gamma} / \zeta(\gamma)$ with N as the maximum observed degree. Estimating γ is then obtained

finding zeros of the derivative of the logarithm of $L(\gamma \mid k)$. As for the exponential-law, $P(k) \approx e^{-\lambda k}/C_1$ where $C_1 = \sum_{k=1}^{\infty} e^{-\lambda k}$ is the normalization constant. The likelihood function is $L(\lambda \mid k) = \prod_{k=1}^{N} e^{-\lambda k}/C_1$ and λ can be obtained by maximizing the logarithm of $L(\lambda \mid k)$ in the same way as for the power-law. Finally, the probability function for the power-law with exponential tail is $P(k) \approx k^{-\alpha} e^{-\beta k}/C_2$ where in this case $C_2 = \sum_{k=1}^{\infty} k^{-\alpha} e^{-\beta k}$. The respective likelihood function is $L(\alpha, \beta \mid k) = \prod_{k=1}^{N} k^{-\alpha} e^{-\beta k}/C_2$. Both, α and β can be obtained again by finding zeros of the derivative of the logarithm of the likelihood function.

2.2 Dependency of the clustering coefficient on the degree

The clustering coefficient of a node v_i measures how many links, n_i, between its neighbors exist in relation to the maximum possible number of such links, i.e. how well its neighborhood is interconnected: $c(v_i) = n_i / k_i(k_i - 1)$. The mean clustering coefficient for vertices with degree k, $C(k)$, is the average clustering coefficient for all nodes with degree k. A network has a hierarchical modular organization if $C(k)$ approximately follows a power-law of kind $C \propto k^{-\omega}$. When $\omega = 1$, modules are arranged in a hierarchical style throughout a network (Ravasz et al., 2002; Barabási and Oltvai, 2004). The power-law approximation of $C(k)$ can be seen in a log-log plot.

2.3 Pairwise disconnectivity index

The pairwise disconnectivity index estimates the importance of a network element for sustaining the connection between pairwise linked nodes (potapov et al., 2008). For a vertex v, it is given by $Dis(v) = 1 - N'/N$ where N' is the number of still pairwise linked vertices when v has been removed from the network and N is the number of all initially pairwise connected vertices: The pairwise disconnectivity index ranges between 0 and 1 where zero indicates that v is not crucial for the connection of any two linked nodes and one means that no two vertices are connected to each other anymore.

3 Results

3.1 Genome-wide mammalian transcription, signaling and metabolic networks

The view of mutually affecting biological processes in a network perspective has greatly facilitated our understanding of the underlying functionality and occasionally complex interrelationships. Such networks may represent specific aspects of regulatory functions in a cell in nearly any detail or just outline the very general processes at a rather discrete level. Here, we consider various mammalian networks at the level of their orthologous abstraction (Choi et al., 2004; Matys et al., 2006; Krull et al., 2006), i.e. pooling the available information for different mammalian species (mostly human, mouse and rat) and thereby neglecting conceivable particularities between them. We have applied this technique to the genome-wide scale, hence focusing on the regulatory potential within the whole genome and not that of its particular parts expressed in various cell types.

We have studied mammalian networks of three different kinds, each of which is represented as a directed graph without double edges. The core network of mammalian transcription factor genes was retrieved from the TRANSFAC® database, release 11.3 (Matys et al., 2006), and the TRANSPATH® database, release 8.3 (Krull et al., 2006). In this network, the nodes represent genes that are coding for transcription factors (TF-genes), other kinds of genes were not considered. The edges respond to the genetic interactions between TF-genes, i.e.

comprise the expression of each TF-gene and the effects of its product, i.e. trans-activation/-repression of the target TF-genes, etc. (Fig. 1). This *transcription* network consists of 279 nodes and 657 edges.
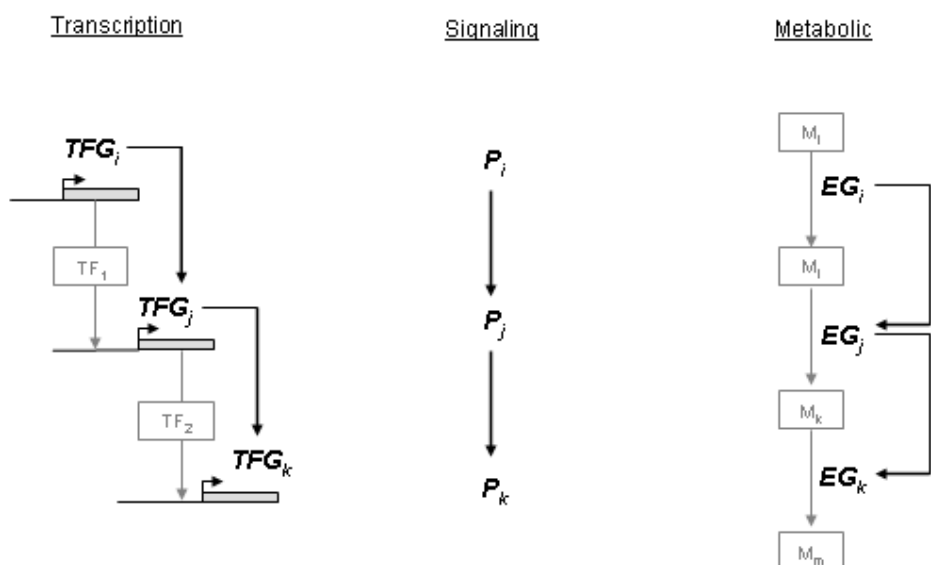

Fig. 1 The semantics of edges in the mammalian networks studied here. *Transcription*: a link from transcription factor gene i (TFG_i) to transcription factor gene j (TFG_j) consists of the expression of TFG_i and the subsequent interaction of transcription factor i (TF_i) with the promoter of transcription factor gene j (TFG_j). *Signaling*: an edge $P_i \rightarrow P_j$ refers to a causal link from protein i to protein j. *Metabolic*: an edge $EG_i \rightarrow EG_j$ represents a link mediated by a metabolite that is both the product and educt of two adjacent metabolic reactions which are catalyzed by enzymes E_i and E_j that are expressed by genes G_i and G_j, respectively.

The TRANSPATH® database (Krull et al., 2006), release 8.3, was also used to reconstruct the mammalian signal transduction network. It shows how various signals are relayed from different receptors to target molecules such as transcription factors and metabolic enzymes. For this, we extracted "semantic" reactions only which focus on the essential components between which information is actively forwarded, and did so at the level of "orthogroups" (Choi et al., 2004). This *signaling* network consists of 1571 nodes and 3425 edges.

The mammalian network of genes encoding enzymes of metabolic reactions was reconstructed from the Ligand section of the KEGG database (Kanehisa et al., 2008) by retrieving all genes encoding metabolic enzyme activity in mammalian (more precisely: human, mouse and rat) systems. Consistently following a genome-centric view, the nodes represent genes coding for metabolic enzymes, and the edge semantics is to forward a metabolite produced by one enzyme to one that consumes it (Fig. 1). This network emphasizes the role of genetically encoded information in metabolic processes and can be viewed as the line graph of the original metabolic graph. We will refer to this one as to *metabolic* network. Its weakly connected part consists of 1793 nodes and 5538 edges.

3.2 The small-world property and presence of modules

The general topological features of these mammalian networks are compiled in Table 1. In the networks the average shortest paths length are with 3.4 (transcription network), 6.2 (signaling network) and 6.4 (metabolic network) fairly small as compared to the respective network size. The mean clustering coefficients of 0.07 in the transcription network, 0.02 in the signaling network and 0.09 in the metabolic network are much larger than the expected values of randomized networks of the same size (Table 1). An increased clustering is indicative for a modular style of network architecture (Barabási and Oltvai, 2004; Ravasz et al., 2002).

Following the criteria of small-world networks stated as increased clustering features and small mean shortest path lengths (Watts and Strogatz, 1998), all three mammalian networks are of small-world.

Table 1 Topological properties of the mammalian regulatory networks

	Transcription	Signaling	Metabolic
Vertices	279	1571	1793
Edges	658	3425	5538
Density	2.36	2.18	3.09
Shortest path length			
Mean	3.4	6.2	6.4
Max	10	18	18
Clustering coefficient			
Mean	0.074	0.021	0.087
Random*	0.016	0.002	0.003
Pairwise disconnectivity index			
Mean	0.012	0.002	0.002
Max	0.211	0.043	0.083

*Expected for random graphs with the same number of vertices and edges.

3.3 Degree distribution does not show scale-freeness

Recent attempts in reassessing the degree distributions of various networks indicate that in many cases, data were fitted prematurely to a power law (PL), $P(k) \propto k^{-\gamma}$, (Amaral et al., 2000; Clauset et al., 2009). The reason for wrongly assigning a power-law can be understood from the way how the degree distribution is usually analyzed (Clauset et al., 2009; Li et al., 2006). In particular, the most frequently applied method - linear regression on a double logarithmic plot - suffers from several major drawbacks and may lead to misleading results (Clauset et al., 2009). In fact, by using other statistical approaches such as the maximum likelihood method, many of the apparently power-law networks were shown to be of single or broad scale (Amaral et al., 2000; Guelzim et al., 2002; Montoya et al., 2006; Khanin and Wit, 2006). Single-scale networks are characterized by a connectivity distribution that decays according to an exponential law (EL), $P(k) \propto e^{-\lambda k}$. In contrast, broad-scale networks display a mixed connectivity distribution that exhibits a power law with an exponential tail (PLET), $P(k) \propto k^{-\alpha} e^{-\beta k}$ (Amaral et al., 2000). Because the exponential decay term $e^{-\beta k}$ overwhelms the power decay term $k^{-\alpha}$ at large k, this distribution is not asymptotic to a power law.

We have analyzed the distributions of the incoming (*k-in*), outgoing (*k-out*) and total (*k-inout*) degrees of the mammalian networks with respect to these three types of models. The fitting parameters were obtained by using a standard maximum likelihood estimators' procedure (see Methods section). Fig. 2 depicts the degree sequences for the incoming, outgoing and total degrees of the networks altogether with the fitted power law, exponential law and the power law with exponential tail. The maximum likelihood estimators for the respective parameters of the models γ (PL), λ (EL), α and β (PLET) are summarized in Table 2. As can be seen from the plots in Fig. 2, pure PL and EL models fit to the observed distributions to different extent, whereas the PLET model always fits well. Only in some cases (all degree distributions of the transcription network and the out-degree distribution of the metabolic network), the pure EL model seems to fit as well as the PLET.
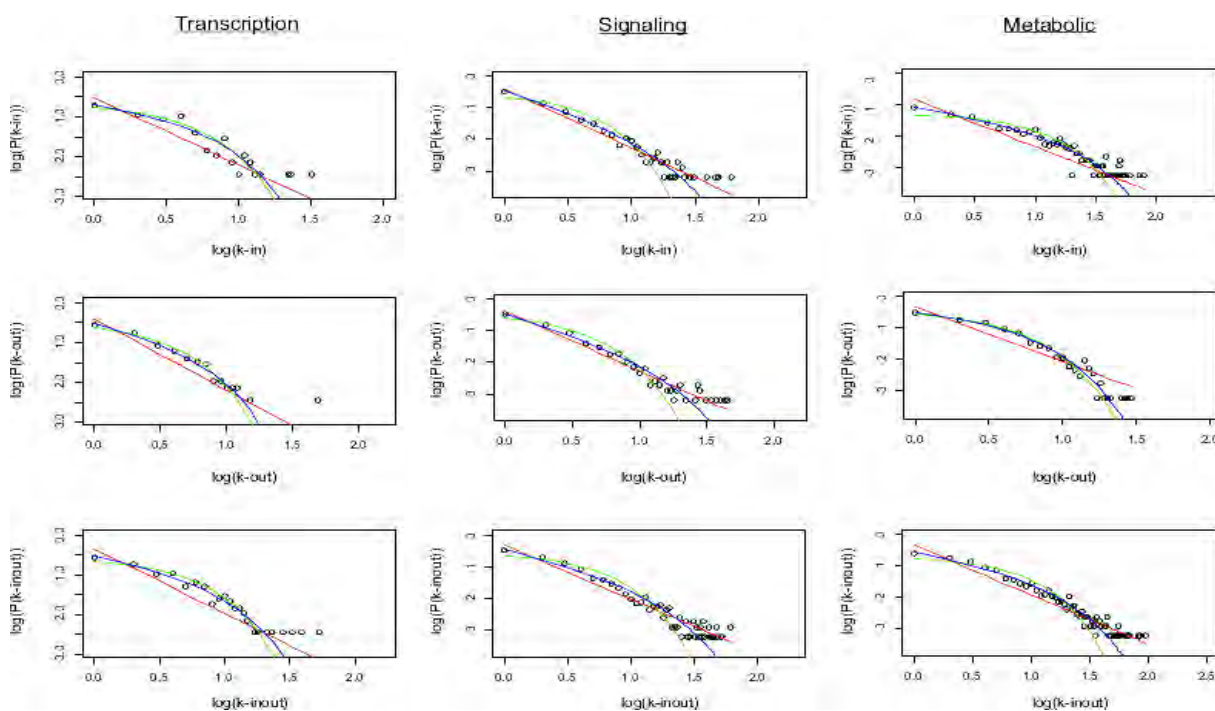
Fig. 2 The *in-*, *out-* and *inout-*degree distributions in the mammalian networks do not confirm pure scale-freeness. The logarithm of the respective degree is denoted on the x-axis and opposed to the logarithm of its observed probability (y-axis). In each of the plots, the red line depicts the fitted power law to the data, the green curve represents the fitted exponential law and the blue curves stands for the fitted power law with exponential tail. The maximum likelihood estimators for the corresponding parameters are specified in the small boxes.

Table 2 Maximum likelihood estimators for the fitted power-law, exponential law and power law with exponential tail models for each of the degree sequences of the transcription, signaling and metabolic networks.

	Power-law	Exponential law	Power law with exponential tail	
	γ	λ	α	β
Transcription				
k-in	1.70	0.24	0.40	0.23
k-out	1.80	0.27	0.70	0.21
k-inout	1.63	0.19	0.56	0.14
Signaling				
k-in	1.86	0.27	1.31	0.09
k-out	1.88	0.29	1.29	0.10
k-inout	1.71	0.21	1.04	0.08
Metabolic				
k-in	1.51	0.12	0.68	0.06
k-out	1.74	0.26	0.33	0.28
k-inout	1.58	0.15	0.78	0.08

   Due to the clear pitfalls of mere visual inspection, we applied a likelihood ratio test which compares the quality of fitness of the three considered models. Table 3 summarizes the differences of the logarithms of the maximum likelihood functions for the three models using their estimated parameters. A positive value indicates that the first of the two models has a higher outcome of its maximum likelihood function and

therefore is the preferred choice compared to the other one, while a negative value denotes the preference of the second model among the two models tested. As assessed above from a visual inspection of the plots in Fig. 2, the likelihood ratio test also reveals a PLET model being superior over the two considered alternatives in fitting the degree distribution of each mammalian network examined (Table 3, the second and third columns). For the transcription network, the preference of a PLET over an EL model is minimal, thereby confirming that the degree distributions of this network can be approximated by the latter. A similar notion can be applied to the outgoing degree distribution in the metabolic network. When just comparing PL and EL models, the power-law model is given preference for the signaling network only, but is nevertheless clearly rejected against the PLET model as a plausible fit for its degree distributions. Hence, none of the networks is scale-free.

Table 3 Likelihood ratio tests for the fitted power-law (**PL**), exponential law (**EL**) and power law with exponential tail (**PLET**) distributions in the mammalian networks. The test is applied to two different models for the *in-*, *out-* and *inout-*degree at a time. A positive (negative) outcome means that the first (second) model is preferred amongst the other one.

	PL vs EL	*PL vs PLET*	*EL vs PLET*
Transcription			
k-in	-32.18	-33.92	-1.74
k-out	-20.34	-25.92	-5.58
k-inout	-45.85	-53.75	-7.90
Signaling			
k-in	76.73	-54.77	-131.50
k-out	63.55	-58.65	-122.21
k-inout	20.70	-149.95	-170.65
Metabolic			
k-in	-101.21	-151.71	-50.50
k-out	-299.37	-308.73	-9.36
k-inout	-157.39	-302.46	-145.07

3.4 Clustering coefficient indicates different architectures of the three mammalian networks

Modular organization is a hallmark of biological systems with each module performing its special functional task (Hartwell et al., 1999; Spirin and Mirny, 2003). Modules are sub-networks of different sizes with an enhanced density of internal links and are thought to be arranged in either a semi-autonomous or hierarchical manner (Ravasz et al., 2002; Barabási and Oltvai, 2004). The semi-autonomous type of organization might be provided by relatively independent groups of interconnected vertices which altogether do not exhibit a dependence of the average clustering coefficient for nodes with degree k, $C(k)$, on the degree k (Ravasz et al., 2002; Barabási and Oltvai, 2004). In contrast, a hierarchical modular organization implies that small groups of vertices assemble hierarchically into increasingly larger groups with communication between the different highly clustered neighborhoods being maintained by a few hubs (Ravasz et al., 2002; Ravasz and Barabási, 2003; Barabási and Oltvai, 2004). Such nested organization of modules is expected to provide a dependence of $C(k)$ on the degree k in a power-law fashion, $C(k) \propto k^{-\omega}$, in a double logarithmic plot (Ravasz et al., 2002; Ravasz and Barabási, 2003; Barabási and Oltvai, 2004). As was shown with artificially designed and highly regular networks constructed by a repeated duplication and integration process of clustered nodes (Ravasz and Barabási, 2003; Barabási and Oltvai, 2004), a slope of -1 indicates a modular architecture in hierarchical style throughout a network. A power-law dependence of the clustering coefficient on the degree ($C(k) \propto k^{-1}$) has been suggested to be the signature of hierarchical networks (Ravasz et al., 2002; Ravasz and Barabási, 2003; Barabási and Oltvai, 2004).

To assess whether the mammalian networks exhibit hierarchical modularity, we checked the dependency of the average clustering coefficient on the degree (Fig. 4). In the case of the transcription network, $C(k)$ clearly depends on k. This dependency can be linearly fitted in a log-log plot and, thus, follows a power-law. However, the overall slope of -0.47 is significantly smaller than the expected -1. This might reflect a relaxed hierarchical setup of modules. On the contrary, $C(k)$ is independent from k in the signaling network and their relation thus strikingly fails to express a power-law behavior in double logarithmic scale. Accordingly, hierarchical modularity cannot be detected in the signaling network and the semi-autonomous type of modular organization is likely to dominate instead. In the metabolic network, there is a dependency of $C(k)$ on k in principle, but the scattering of values in the log-log plot (Fig. 3) does not allow a plausible approximation by a power-law. Probably, hierarchical and semi-autonomous types of modularity co-exist in this network.
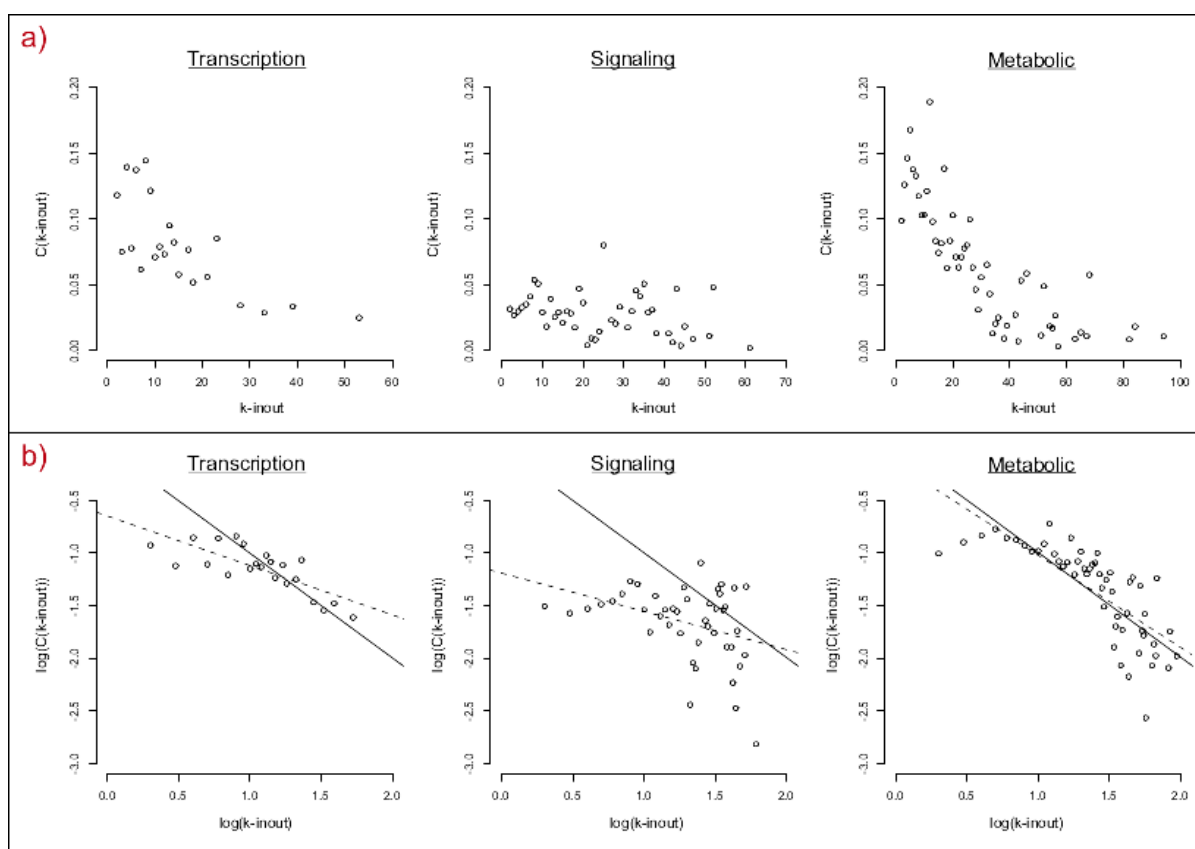

Fig. 3 The dependency of the mean clustering coefficient $C(k)$ on the degree k. (a): Relationships between $C(k)$ and k in non-logarithmic scale. (b): For each of the mammalian networks, the logarithm of the mean clustering coefficient of all vertices with *inout*-degree k is plotted against the logarithm of k. The diagonal solid line in each figure has slope 1, following $C(k) \propto k^{-1}$, which corresponds to the case of throughout hierarchical modularity. The dashed line represent the best linear fit.
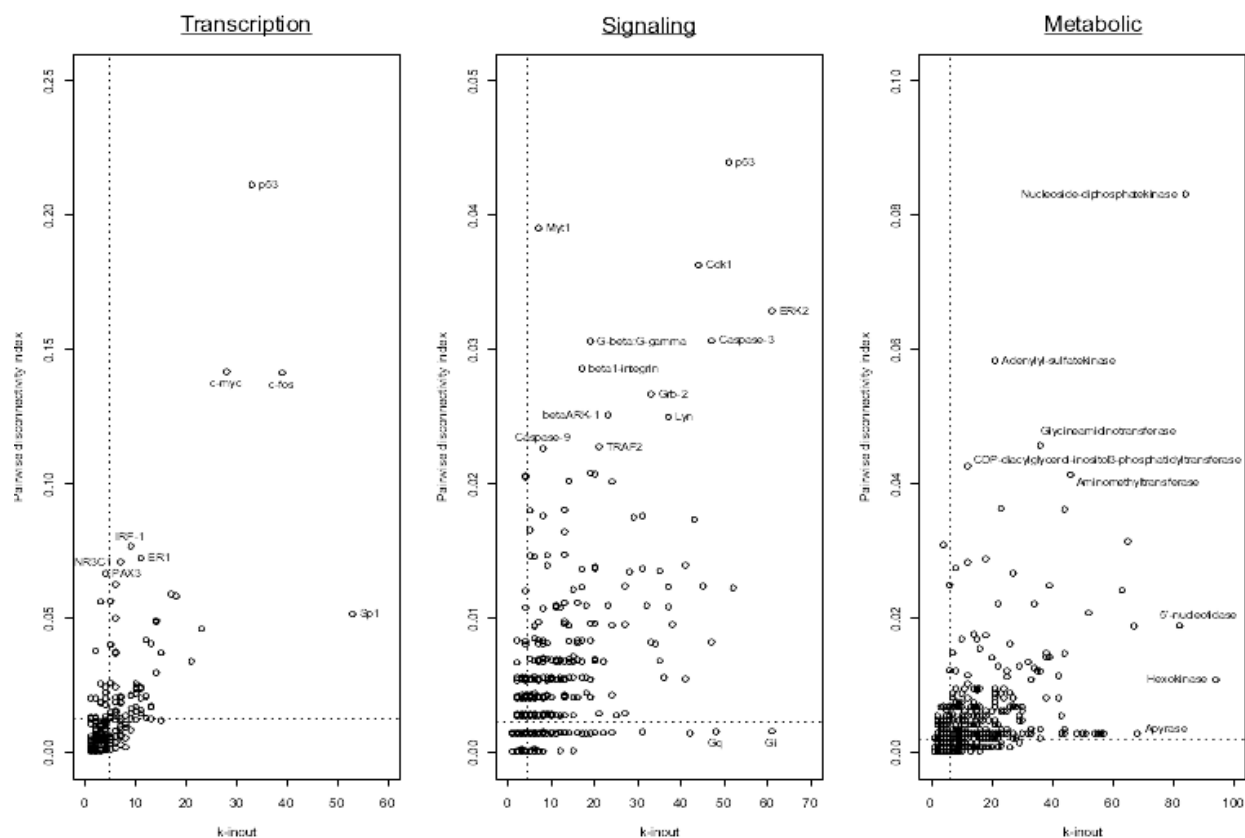
Fig. 4 The pairwise disconnectivity index of individual vertices versus their *inout*-degrees. Although the topological significance of a vertex (x-axis) positively correlates with the degree (y-axis), this trend is not straightforward. Hub nodes are neither generally crucial for the connectedness of the networks nor are low or mid-range degree nodes constantly playing a minor role in this context. Instead, a strong effect on the connectedness that is far in excess of the average pairwise disconnectivity indices (dotted vertical lines) occurs for nearly all kinds of degrees. The mean pairwise disconnectivity index and the mean *inout*-degree of all vertices in a network are indicated by the horizontal and vertical dotted lines, respectively.

3.5 Network robustness and the importance of hub nodes

Next to a modular organization, biological systems fundamentally feature a remarkable robustness (Kitano, 2004). It has been observed that they can generally withstand perturbations, but are sensitive against targeted attacks on the hub nodes (Albert et al., 1999; Albert et al., 2000; Barabási and Oltvai, 2004; Crucitti et al., 2004). In this regard, we have examined the three mammalian networks to their tolerance against single node knockouts using the methodology introduced in (Potapov et al., 2008) – the pairwise disconnectivity index (PDI). This approach is based on the systematic removal of individual nodes and computing the number of ordered pairs of vertices that become disconnected due to the erasing of a given vertex from the network. The PDI of node *v* characterizes the topological significance of *v* by considering those direct and indirect connections in the network that critically depend on the presence of *v* (Potapov et al., 2008). The method has been shown to be a well-suited measure for detecting key nodes and groups of nodes in regulatory networks (Potapov et al., 2008; Goemann et al., 2009a; Goemann et al., 2009b).

By applying the PDI to the mammalian networks, we found that random removal of any individual node has only marginal impact on the connectedness of these networks (Fig. 4). The respective mean PDI values are 0.012 (transcription), 0.0023 (signaling) and 0.0019 (metabolic), i.e. deleting a node expectedly disrupts the connection of just 1.2%, 0.23% and 0.19% of the existing paths. However, a small number of nodes are characterized by a PDI value that significantly exceeds the average PDI values (up to 0.21), thereby rendering

the networks vulnerable upon a targeted removal (Fig. 4). These results clearly evidence that the networks are not centralized: none of nodes globally controls the integrity of each network.

Interestingly, these key nodes are not necessarily hubs. As Fig. 4 shows, hubs associated with a high PDI value surely exist in the mammalian networks, e.g. *c-fos, c-myc* and *p53* in the transcription network. However, there are also hubs the topological significance of which ranges within the same scale as selected low or mid-range degree vertices (e.g. *Sp1* in the transcription network). In particular within the signaling and metabolic networks many examples were found where hubs have just a small or nearly no impact at all (e.g. *Gi* in the signaling and *apyrase* in the metabolic network). On the other side, a strong effect on the connectedness of all three networks was observed for many nodes with much smaller degrees like *IRF-1* in the transcription and *Myt1* in the signaling network. Consequently, and in agreement with (Potapov et al., 2008), no simple and straightforward relationship between a node's importance for the connectedness of a network and its degree can be detected. The degree alone does not seem to be the decisive criteria for the topological importance of a node.

4 Discussion

Regulation of cellular processes in higher eukaryotes is characterized by an outstanding complexity. Diverse programs, each involving a large number of molecular entities and their interactions, run in different time scales in an apparently independent manner although being closely interrelated in reality. They generate and forward input-dependent outputs to each other, thereby spotlighting their inner machineries, which enable them to react specifically on a certain composition of inputs (signals, hormones, nutrients, etc.). Much effort has therefore been put into understanding the configurations of the basic regulatory mechanisms.

In this context, several architectural features were detected and advocated as generic attributes of biological regulatory networks: Amongst others, these are (*i*) scale-freeness given by a power-law degree distribution (Barabási, 2009), (*ii*) hierarchical modularity (Ravasz and Barabási, 2003), and (iii) a robust yet fragile design with hubs as the key nodes (Albert et al., 1999, Li et al., 2006). Consequently, very different functional systems seem to share fundamental properties, notwithstanding that some of the concepts, namely the prevalence of power-law degree distributions and the importance of motifs (Milo et al., 2002), are controversially discussed (Goemann et al., 2009a; Goemann et al., 2009b; Guelzim et al., 2002; Khanin and Wit, 2006; Konagurthu and Lesk, 2008; Lima-Mendez and van Helden, 2009).

4.1 Mammalian networks are not scale-free

With regard to the concepts mentioned above, we have characterized the topologies of three mammalian networks – transcription, signaling and metabolic – to identify both their global traits and the rather specific properties of their individual vertices. None of the degree distributions of these networks could be convincingly approximated by a power-law, which coincides with the findings of recent studies that have (re-) analyzed the degree sequences of several other biological networks (Amaral et al., 2000; Montoya et al., 2006; Guelzim et al., 2002; Khanin and Wit, 2006; Lima-Mendez and van Helden, 2009). Rather, the best fits were obtained by a power-law with an exponential tail (PLET), which associates power-law behaviour for low-degree nodes fading progressively to exponential law as the degree increases (Amaral et al., 2000; Clauset et al., 2009; Khanin and Wit, 2006). This signifies that none of the abstracted mammalian networks, as a whole, is scale-free. However, the lack of scale-freeness does not entail the absence of a hub-centric organization of these networks. Despite the notion of network hubs has originally been associated with a power-law degree distribution, it may equally apply to networks with other types of organization, as in the case of the inspected mammalian networks.

For the transcription network, we found that its degree distributions can also be very well approximated by an exponential law. Such a distribution was reported for the *in*-degree of the transcription networks of yeast (Guelzim et al., 2002) and bacteria (Teichmann and Babu, 2004; Balaji et al., 2007). In contrast, the *out*-degree distributions for these networks do not follow an exponential law (Guelzim et al., 2002; Teichmann and Babu, 2004; Balaji et al., 2007), which differs from our findings for the mammalian transcription network. These differences may reflect some peculiarities of transcriptional regulation in mammals as compared with that in bacteria and yeast. Besides, the mammalian transcription network studied here did not include any non-TF-genes, while the bacterial and yeast networks (Guelzim et al., 2002; Teichmann and Babu, 2004; Balaji et al., 2007) included both TF-genes and non-TF-genes. Conceivably, the core transcription network comprising only TF genes and the extended transcription network, which also includes non-TF target genes, may be differently organized. Preliminary studies of our group support this hypothesis (M. Haubrock et al., in preparation).

The signaling network is the only one where in the direct comparison of the two one-parameter models (PL and EL), the power-law (PL) model is superior to the exponential law (EL) model. However, even in this case the hypothesis that this network is scale-free is rejected when comparing with the PLET model. For the metabolic network our data neither supports a power-law connectivity distribution of metabolites (Jeong et al., 2000; Wagner and Fell, 2001) nor a coinciding in- and out-degree distribution of metabolites (Jeong et al., 2000). Its out-degree distribution could be well fitted to an exponential law. That is in accordance with the observation that the connectivity distribution of metabolites in separate functional modules is exponential rather than a power law (Tanaka, 2005).

The absence of pure scale-freeness in the mammalian networks is rather important for understanding the possible mechanism of their evolutionary development. The developmental model based on a preferential attachment of newly emerging vertices to hubs has been proposed to explain how the power-law degree distribution can emerge in biological networks (Barabasi and Albert, 1999; Barabási and Oltvai, 2004; Albert, 2005). Since the mammalian networks significantly deviate from a power law, the contribution of other developmental mechanisms should be considered as well to explain the complexity of the networks. We suggest that the subsets of genes and relationships that constitute the mammalian networks have co-evolved very differently and through multiple mechanisms.

4.2 The mammalian networks differ in their modular organization

The three mammalian networks differ in the dependencies of the mean clustering coefficient $C(k)$ on the degree k. In both the transcription and the metabolic network, $C(k)$ decreases with the degree. With regard to the criteria of hierarchical modularity (Ravasz et al., 2002; Barabási and Oltvai, 2004), these two networks, but not the signaling one, are therefore expected to contain hierarchically organized modules. This implies that while the low-degree vertices are part of highly cohesive, densely interlinked clusters, the high-degree vertices are not, as their neighbors have a smaller chance of linking to each other (Ravasz et al., 2002; Barabási and Oltvai, 2004). Hence, the high-degree vertices may play the role of bridging the many small communities of clusters into larger and hierarchically organized parts.

In the case of the transcription network, the linear dependency of $C(k)$ on k indicates a largely hierarchical setup of modules in this network. Since the overall slope deviates from the standard -1, this constitution may not be a pure, but rather an "idealized" one, which in reality may be diluted by other components.

In contrast, the mammalian signaling network appears to be predominantly anti-hierarchical: most of its modules relate to one another in a semi-autonomous fashion. This means that signal processing occurs predominantly *within* individual modules and signal propagation *between* different modules through defined interfaces. Since signal transduction is mostly conveyed by protein-protein interactions, this network structure reflects the fact that its edges depend on the presence of one or few usually highly specific binding sites in

each of the constituents (nodes). It ensures the necessary specificity of signal transduction from distinct receptors to distinct targets, such as transcription factors, metabolic enzymes etc., and at the same time efficient re-use of some modules for different purposes. To the best of our knowledge, this particular feature of the mammalian signaling network has not yet been reported, although this observation is in agreement with the general functional role of signal transduction systems.

The more complex dependency of $C(k)$ on k in the metabolic network, i.e. the clear lack of a consistent power-law dependence, may suggest that parts of the network are organized in hierarchical modules, while other parts are constituted by semi-autonomous modules. Such a mixed architectural design is likely to refer well to the division of metabolism into anabolism, catabolism and central metabolism which are subdivided further into multiple incoming and outgoing pathways that label the main functional routes (e.g. glycolysis, tricarboxylic acid cycle) and join at different places to form an interconnected network (Fell, 2007). The co-existence of various types of modular organization may provide an additional advantage: networks with such a fuzzy community structure are expected to be more efficient in executing the represented processes than those with a pronounced community structure (Danon et al., 2008).

4.3 The three mammalian networks differ in their robustness

By performing a knockout analysis of individual nodes, we have shown that the connectivity of the mammalian networks remains almost unaffected when randomly choosing the node. However, targeting one out of a tiny fraction of selected nodes significantly perturbs the associated network. Some of these particularly important nodes are hubs, but some others are mid-range and low-degree vertices. In accordance with our previous results (Potapov et al., 2008), there is no obvious relationship between the topological significance of a node as measured by its PDI, and its degree. Rather, this significance is determined by the role a node plays for the global connectivity of the whole network, while the degree is only a local property.

In particular the transcription network is characterized by a set of nodes that convey high vulnerability. This, together with the very small average shortest path length and the characteristic dependence of the clustering coefficient on the node degree discussed above, characterizes the architectural design of the transcription network as the most compact and unified among the three networks studied here. The network therefore is hierarchically centralized around a limited set of nodes. This peculiarity of the transcription network is reasonable when considering that in higher eukaryotes the transcription of a gene is generally controlled by a combination of several transcription factors that act together and in a cooperative fashion.

Removing a single node exerted only moderate effects on the pairwise connectivity within the signaling and the metabolic network indicating comparable robustness for both these networks. The high robustness of protein-to-protein connections in the signaling network is particularly remarkable because this network is the most sparsely connected one among the three networks studied here. It means that connections in the signaling network seem to be backed up by many alternative paths, thus reducing the dependency on single nodes. The semi-autonomous modular organization of the signaling network ensures that a local perturbation is restricted to the affected module, leaving the remainder of the network largely operable. These features of the signaling network also respond well to the functional role of this network in relaying regulatory information via signaling pathways. Note that the observed redundancy refers to a genome-wide signaling network. Since various cell types express only particular parts of this network, the actual redundancy of cell-specific networks might be greatly relaxed.

4.4 Critical survey

Obviously, the information used to create these networks, as well as our whole current knowledge about these mammalian systems is not complete. We cannot exclude that some of the numeric values in our estimates will not survive while completing the knowledge. In particular, this might relate to the transcription network, as the

most incomplete one. But, nevertheless, we believe that the ascertained features of the inspected networks constitute the main trends of the underlying yet uncharted entire systems.

5 Conclusions

The topological properties of the investigated networks reveal distinct architectures. The transcriptional network exhibits a hierarchical modularity, whereas the signaling network is mainly comprised of semi-autonomous modules. The metabolic network, in contrast, seems to be constituted of a more complex mixture of substructures. In any case, high-PDI nodes are considered to play the major role in interlinking the different modules. We conclude that the subsets of genes and relationships that constitute these networks have co-evolved very differently and through multiple mechanisms.

Acknowledgements

References

Agrawal H. 2002. Extreme self-organization in networks constructed from gene expression data. Physical Review Letters, 89(26): 268702

Albert R, Jeong H, Barabási A-L. 2000. Error and attack tolerance of complex networks. Nature, 406(6794): 378-382

Albert R. 2005. Scale-free networks in cell biology. Journal of Cell Science, 118(Pt 21): 4947-4957

Amaral LAN, Scala A, Barthélémy M, Stanley HE. 2000. Classes of small-world networks. Proceedings of the National Academy of Sciences of USA, 97(21): 11149–11152

Balaji S, Babu MM, Aravind L. 2007. Interplay between network structures, regulatory modes and sensing mechanisms of transcription factors in the transcriptional regulatory network of *E. coli*. Journal of Molecular Biology, 372(4): 1108–1122

Barabási A-L and Albert R. 1999. Emergence of scaling in random networks. Science, 286(5439): 509-512

Barabási A-L, Oltvai ZN. 2004. Network biology: understanding the cell's functional organization. Nature Reviews Genetics, 5(2): 101-113

Barabási AL. 2009. Scale-free networks: a decade and beyond. Science, 325(5939): 412-413

Bhan A, Galas DJ, Dewey TG. 2002. A duplication growth model of gene expression networks. Bioinformatics, 18(11): 1486-1493

Bhardwaj N, Yan KK, Gerstein MB. 2010. Analysis of diverse regulatory networks in a hierarchical context shows consistent tendencies for collaboration in the middle levels. Proceedings of the National Academy of Sciences of USA, 107(15): 6841–6846

Carter SL, Brechbuhler CM, Griffin M, et al. 2004. Gene expression network topology provides a framework for molecular characterization of cellular state. Bioinformatics, 20(14): 2242-2250

Choi C, Crass T, Kel A, et al. 2004. Consistent re-modeling of signaling pathways and its implementation in the TRANSPATH database. Genome Informatics, 15(2): 244-254

Clauset A, Shalizi C, Newman MEJ. 2009. Power-law distributions in empirical data. SIAM Review, 51(4): 661-703

Crucitti P, Latora V, Marchiori M, et al. 2004. Error and attack tolerance of complex networks. Physica A: Statistical Mechanics and its Applications, 340(1-3): 388-394

Danon L, Arenas A, Díaz-Guilera A. 2008. Impact of community structure on information transfer. Physical

Review E, 77(3 Pt 2): 036103

Dorogovtsev SN, Mendes JFF. 2002. Evolution of networks. Advances in Physics, 51: 1079-1187

Featherstone DE, Broadie K. 2002. Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. Bioessays, 24(3): 267-274

Fell DA. 2007. Metabolic networks. In: Biological Networks. Complex Systems and Interdisciplinary Science (Vol. 3) (Kopos F, ed). World Scientific Publishing Co., New Jersey, USA

Giot L, Bader JS, Brouwer C, et al. 2003. A protein interaction map of *Drosophila melanogaster*. Science, 302(5651): 1727-1736

Goemann B, Wingender E, Potapov AP. 2009a. An approach to evaluate the topological significance of motifs and other patterns in regulatory networks. BMC Systems Biology, 3: 53

Goemann B, Potapov AP, Ante M, et al. 2009b. Comparative analysis of topological patterns in different mammalian networks. Genome Inf Ser, 23(1): 32-45

Guelzim N, Bottani S, Bourgine P, et al. 2002. Topological and causal structure of the yeast transcriptional regulatory network. Nature Genetics, 31(1): 60-63

Hartwell LH, Hopfield JJ, Leibler S, Murray AW. 1999. From molecular to modular cell biology. Nature, 402(6761 Suppl): C47-C52

Ibrahim SS, Eldeeb MAR, Rady MAH. 2011. The role of protein interaction domains in the human cancer network. Network Biology, 1(1): 59-71

Jeong H, Mason SP, Barabási A-L, et al. 2001. Lethality and centrality in protein networks. Nature, 411(6833): 41-42

Jeong H, Tombor B, Albert R, et al. 2000. The large-scale organization of metabolic networks. Nature, 407(6804): 651-654

Kanehisa M, Araki M, Goto S, et al. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Research, 36(Database issue): D480-D484

Khanin R, Wit E. 2006. How scale-free are biological networks? Journal of Computational Biology, 13(3): 810-818

Kitano H. 2004. Biological robustness. Nature Reviews Genetics, 5(11): 826-837

Konagurthu A, Lesk A. 2008. On the origin of distribution patterns of motifs in biological networks. BMC Systems Biology, 2: 73-81

Krull M, Pistor S, Voss N, et al. 2006. TRANSPATH®: An information resource for storing and visualizing signaling pathways and their pathological aberrations. Nucleic Acids Research, 34(Database issue): D546-D551

Li L, Alderson D, Doyle J, et al. 2006. Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications. Internet Mathematics, 2(4): 1-39

Lima-Mendez G, van Helden J. 2009. The powerful law of the power law and other myths in network biology. Molecular Biosystems, 5(12): 1482-1493

Luscombe NM, Babu MM, Yu H, et al. 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. Nature, 431(7006): 308-312

Ma H, Zeng AP. 2003. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics, 19(2): 270-277

Ma'ayan A, Jenkins SL, Neves S, et al. 2005. Formation of regulatory patterns during signal propagation in a mammalian cellular network. Science, 309(5737): 1078-1083

Martínez-Antonio A. 2011. *Escherichia coli* transcriptional regulatory network. Network Biology, 1(1): 21-33

Maslov S, Sneppen K. 2002. Specificity and stability in topology of protein networks. Science, 296(5569):

910-913

Matys V, Kel-Margoulis OV, Fricke E, et al. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Research, 34(Database issue): D108-D110

Milo R, Shen-Orr S, Itzkovitz S, et al. 2002. Network motifs: Simple building blocks of complex networks. Science, 298(5594): 824-827

Montoya JM, Pimm SL, Solé RV. 2006. Ecological networks and their fragility. Nature, 442(7100): 259-264

Newman MEJ. 2003. The structure and function of complex networks. SIAM Review, 45: 167-256

Papin JA, Hunter T, Palsson BO, et al. 2005. Reconstruction of cellular signaling networks and analysis of their properties. Nature Reviews Molecular Cell Biology, 6(2): 99-111

Papin JA, Palsson BO. 2004. Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. Journal of Theoretical Biology, 227(2): 283-297

Potapov AP, Goemann B, Wingender E. 2008. The pairwise disconnectivity index as a new metric for the topological analysis of regulatory networks. BMC Bioinformatics, 9: 227

Potapov AP, Voss N, Sasse N, et al. 2005. Topology of mammalian transcription networks. Genome Inform, 16(2): 270-278

Ravasz E, Barabási A-L. 2003. Hierarchical organization in complex networks. Physical Review E, 67(2 Pt 2): 026112

Ravasz E, Somera AL, Mongru DA, et al. 2002. Hierarchical organization of modularity in metabolic networks. Science, 297(5586): 1551-1555

Spirin V, Mirny LA. 2003. Protein complexes and functional modules in molecular networks. Proceedings of the National Academy of Sciences of USA, 100(21): 12123-12128

Tacutu R, Budovsky A, Yanai H, et al. 2011. Immunoregulatory network and cancer-associated genes: molecular links and relevance to aging. Network Biology, 1(2): 112-120

Tanaka R. 2005. Scale-rich metabolic networks. Physical Reviews Letters, 94: 168101

Teichmann SA, Babu MM. 2004. Gene regulatory network growth by duplication. Nature Genetics, 36(5): 492–496

Wagner A, Fell DA. 2001. The small world inside large metabolic networks. Proceedings of the Royal Society of London B, 268(1478): 1803-1810

Wagner A. 2001. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. Molecular Biology and Evolution, 18(7): 1283-1292

Watts DJ, Strogatz SH. 1998. Collective dynamics of 'small-world' networks. Nature, 393(6684): 409-410

Yook SH, Oltvai ZN, Barabási A-L. 2004. Functional and topological characterization of protein interaction networks. Proteomics, 4(4): 928-942

Zhang WJ. 2011. Constructing ecological interaction networks by correlation analysis: hints from community sampling. Network Biology, 1(2): 81-98