

How scale-free are biological networks

Raya Khanin and Ernst Wit

Department of Statistics, University of Glasgow

15 University Gardens, Glasgow G12 8QW, U.K.

1. Introduction

Cellular biological interactions have been studied on several different levels. It is possible to study interactions in the metabolome, or on a protein-level or on the level of gene interactions (Figure 1). Since the information on these interactions is rarely quantitative, they have typically been represented via edges in a graph, where the nodes stand for either genes or gene products, such as proteins or metabolites. Collection of these cellular interactions yield a network whose general structure (or topology) has been a subject of recent intensive studies. Whatever level has been selected, it seems that several conclusions are robust to the choice of data: (i) there is a short path from any node to another node (*small world property*); (ii) there are many nodes with few connections and a few nodes with many connections (*hubs*) and (iii) hubs are enriched with essential/lethal nodes (*centrality & lethality principle* coined by Jeong et al., 2001)

An important measure of the network topology is the distribution of the number of connections per node (*connectivity distribution*) (Barabasi and Oltvai, 2004). Metabolic networks (Jeong et al. 2000; Fell and Wagner, 2000; Ma and Zeng, 2003), protein-protein interactions networks (Jeong et al. 2000; Maslov and Sneppen, 2002), protein domain networks (Rzhetsky and Gomez, 2001; Wuchty, 2001), gene interactions (Tong et al. 2004) and gene expression networks (Featherstone and Broadie, 2002; Agrawal, 2002; Bergmann et al. 2004; van Noort et al. 2004) have been reported to exhibit *scale-free* behaviour based on the distribution of the connectivities of the network nodes. Although networks are commonly referred as being scale-free, it is their connectivity (also called degree) distribution that is considered to be scale-free.

The search for biological properties that exhibit scale-free behaviour extends beyond networks of interactions, with examples from the distribution of gene expression levels (Ueda et al. 2004) and the spot intensities (Hoyle et al. 2002) on microarrays to the frequency of occurrence of generalized parts in genomes of different organisms (Luscombe et al. 2002). Thus, the concept of scale-free network has emerged as a powerful unifying paradigm in the study of complex networks in biology, physics and social sciences (Barabasi and Oltvai, 2004).

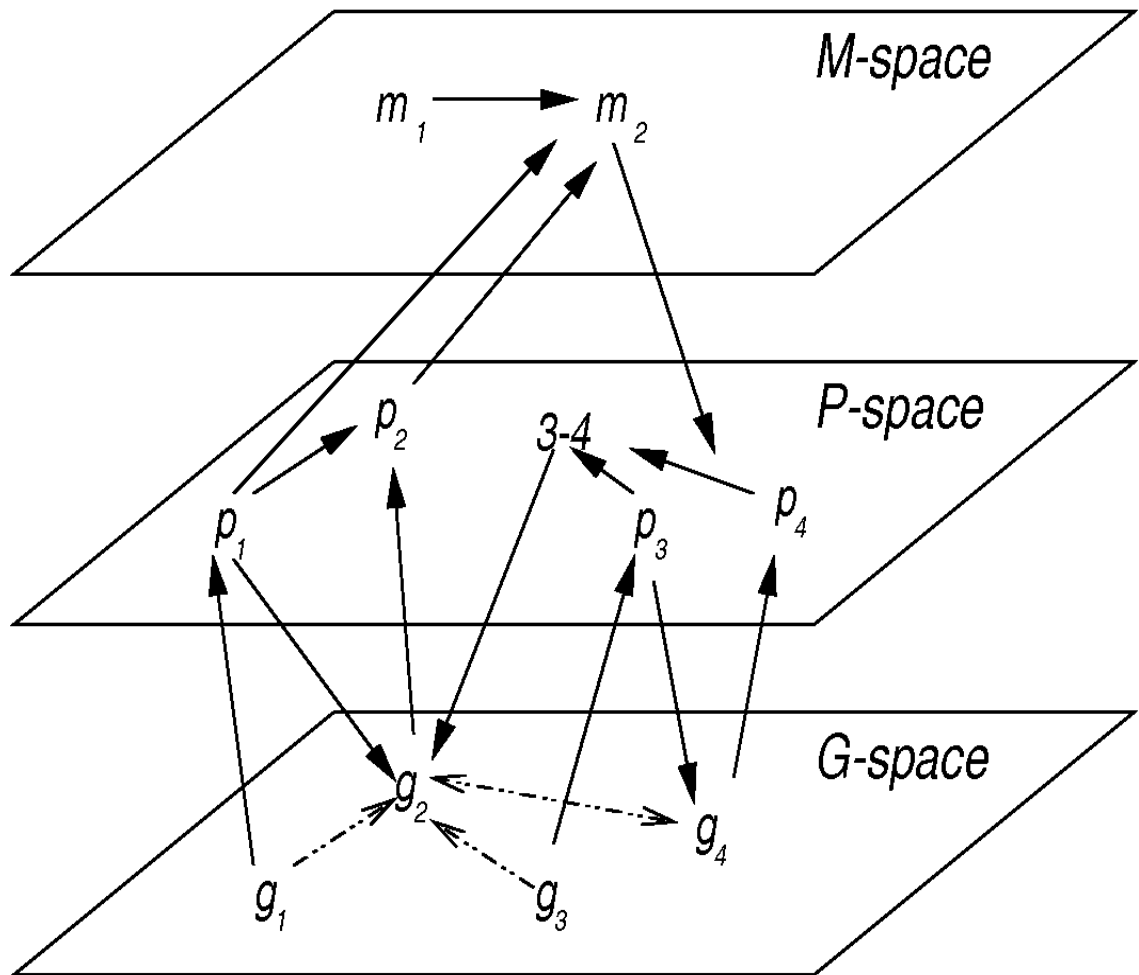


Figure 1. This representation of cellular biological interactions gives a general overview of how biological networks are represented on different levels. It is an abstract representation of a highly complex network of interactions between genes, proteins and metabolites. These networks are commonly studied independently of each other, in their own spaces, while they are obviously closely interconnected and interrelated. Here G-space stands for a space of gene interactions, P-space is a space of protein interactions and M-space is a space of interactions between metabolites. Solid arrows represent direct causal interactions, whereas the dotted arrows are indirect gene interactions, which occur via intermediate causal interaction(s). A widely used graph approach represents networks as graphs wherein a node stands for a gene, protein or metabolite, and an edge represents an interaction between the two nodes. All three types of considered networks have been previously studied using a graph-approach and they have all been shown to share certain characteristics, such as small-world property and existence of hubs with a high proportion of essential/lethal nodes.

Why is the interest in scale-freeness of biological networks has been so great? Scale-free networks possess a number of intriguing properties. Firstly, the presence of several hubs in the biological networks and a large number of nodes

with a few connections has been cited as the most characteristic feature of a scale-free network. Secondly, scale-free networks belong to the class of small world networks (Amaral et al. 2000). This small-world property of networks allows fast communication between different nodes. Scale-free networks have also been reported to be robust to random breakdowns (Albert et al. 2000). All these properties are vitally important for biological networks. However, they are not unique to scale-free networks. One can come up with other functional forms that will exhibit the small-world property as well as the enrichment of the network with highly connected nodes.

A unique property that scale-free systems possess is their invariance to changes in scale. The term 'scale-free' refers to a system defined by a functional form $f(x)$ that remains unchanged within a multiplicative factor under a rescaling of the independent variable x . Effectively, this means power-law forms, since these are the only solutions to $f(ax) = bf(x)$ for all $x \in \mathbb{N}$ (Newman, 2003). The scale-invariance property is often interpreted as the *self-similarity*. Any part of the scale-free network is stochastically similar to the whole network and parameters are assumed to be independent of the system size (Jeong et al. 2000). Other mathematical laws that might fit to describe similar qualitative properties of the network degree distribution will not satisfy an important condition of the scale invariance. Therefore, a network is defined as scale-free if a randomly picked node has k connections with other nodes with a probability that follows a power-law $p(k) \sim k^{-\gamma}$, where γ is the power-law exponent.

Since other than power-law mathematical forms do not possess the scale-invariance (self-similarity) property, the exact form of the degree distribution of biological networks has important implications for the biological conclusions that can and cannot be drawn from the actual topology of the architecture of the network.

The very question on whether and how the topology of biological networks can provide insights into the biology is still debated (see review Alm and Arkin, 2003; Huang, 2004). It is widely believed that unravelling the network topology will lead to understanding the design principles of biological networks and therefore provide some clues into the dynamical evolutionary processes that generated these networks (Jeong et al, 2000).

In pursuit of further understanding of these processes, formal evolutionary models that lead to generation of scale-free network architecture have already been proposed. It is inevitable that such models are oversimplifications of reality. However, the underlying assumptions of these evolutionary models are important as they can provide clues to our understanding of evolution. Several plausible evolutionary mechanisms have been suggested, among them the process of duplication, whereby a newly created node inherits all or part of the connections of the parent node; a divergent "big bang" model with evolutionary drift (Dokholyan et al, 2002); and a birth-death-innovation model (Karev et al, 2002). The duplication models (Bhan et al, 2002; Rzhetsky and Gomez, 2001; van Noort

et al, 2004) have been shown to result in scale-free networks, while plausible evolutionary phenomena such as evolutionary drift has been shown to contradict the power-law (Przytycka and Yu, 2004).

Often, the underlying principles and assumptions of evolutionary models are adjusted so that they yield the scale-free topology of the network. For example, the scale-free network can appear from the model of genome evolution if there is a precise balance between the rates of birth, death and innovation (Karev et al, 2002). It is therefore vitally important to determine the topology of the biological networks correctly and not to rely on oversimplified assumptions of the scale-freeness of their connectivity distribution. Potentially, there are other biological implications of the network topology but those are still debated (Alm and Arkin, 2003; Huang, 2004; Wolf et al, 2002).

The goal of this paper is to establish whether indeed the biological networks are scale-free. By applying formal tools of statistical analysis, we here show that the scale-free assumption does not hold for any of the networks that we have analyzed.

2. Analysis.

A widely used method for finding an indication of a scale-free network, has been to fit a straight line, to the connectivity relative frequency distribution, $p(k) \sim k^{-\gamma}$, $k \geq 1$ on a log-log plot, where k stands for the number of connections (connectivity) of a node (Jeong et al. 2000; Ma and Zeng, 2003; Bergmann et al. 2004; Guelzim et al. 2002; Aloy and Russell, 2004). It has become tempting to describe any decreasing experimental data on a number of connections in a biological network on a log-log plot by a linear fit. An example of fitting a straight line $\log(N(k)) = \alpha - \gamma \log(k)$ to two interaction datasets can be seen in Figure 2, where $N(k)$ stands for the observed connectivity frequency distribution. It is doubtful that a linear fit here can be taken as a good indicator of the power-law distribution. Notice that the points in the plot represent unequal numbers of observations, which is ignored by the linear fit. Moreover, fitting a straight line doesn't necessarily make the points follow it. More formal tools for finding out whether the connectivity distribution is indeed described by the power-law are therefore required.

As a critical evaluation of the scale-free property of biological networks, we performed a statistical re-analysis of datasets of interactions of biological networks that have been reported to be scale-free. Given the importance of the scale-free property, we estimate the exponent, γ , by fitting the power-law to the data using the maximum-likelihood method and perform a goodness-of-fit test to determine whether the data is indeed being drawn from the power-law distribution.

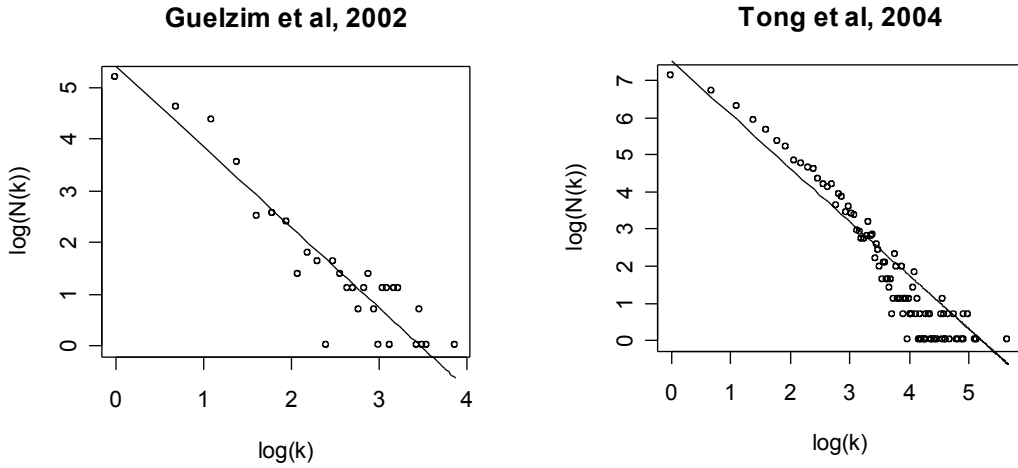


Figure 2. Examples of fitting straight lines to the connectivities of two interaction datasets.

The parameter γ in the power-law density, $p(k) = \frac{k^{-\gamma}}{\zeta(\gamma)}$, where $\zeta(\gamma)$ is the Riemann zeta-function (truncated for $\gamma \leq 1$) and $k \geq 1$, has been estimated by maximum likelihood for each of the networks. The number of connections (connectivity), x_i , for a node i is often obtained from experimental data. In a large network the number of connections can be assumed to be approximately independent. (In the Appendix we show that the assumption of independence of connectivities of all nodes in the network can be weakened by assuming independence of connectivities of nodes in a smaller subnetwork and this does not affect the main conclusion of the paper.)

As a result, the likelihood function can be written as $L(\gamma | x) = \prod_{i=1}^N x_i^{-\gamma} / \zeta(\gamma)$, where N is the maximum observed connectivity, i.e. potentially the total number of genes in the network minus one. The log-likelihood $l(\gamma | x) = -\gamma \sum_{i=1}^N \log x_i - N \log \zeta(\gamma)$ is maximized by finding zeros of its derivative using the Newton-Raphston method. In order to test whether the connectivity data from a particular network can be considered scale-free, we consider a chi-squared statistic, $T = \sum_{k=1}^{k^*} (O_k - E_k)^2 / E_k \sim \chi_{k^*-2}^2$. Here O_k are the observed values of connectivities from the data and E_k are the values estimated from the power-law with γ estimated by maximum likelihood as described above. Using standard statistical procedure, we pool for connectivity values over k^* , for which the expected number of connections is less than 5. As a result, the

chi-squared statistic is approximately chi-squared distributed with $k^* - 2$ degrees of freedom, if the data come truly from the power-law distribution. As a result, the P -value for each of the networks can be calculated by the exceedence probability of a chi-squared distribution.

For the exponentially truncated power-law, the density is given by $p(k) = \frac{k^{-\gamma} \exp(-k/k_c)}{C_1(\gamma, k_c)}$, where $C_1(\gamma, k_c) = \sum_{j=1}^N j^{-\gamma} \exp(-j/k_c)$, γ is the power-law exponent and k_c is the cut-off parameter. Likelihood can be written as $L(\gamma | x) = \prod_{i=1}^N x_i^{-\gamma} \exp(-x_i/k_c) / C_1(\gamma, k_c)$ and the log-likelihood $l(\gamma | x) = -\gamma \sum_{i=1}^N \log x_i - 1/k_c \sum_{i=1}^N x_i - N \log C_1(\gamma, k_c)$ is maximized with respect to two parameters γ and k_c by finding zeros of the derivative using the Newton-Raphson method. Corresponding P -values are calculated similarly to the power-law by considering a chi-squared statistic.

R-code used for calculations and some instructions can be found on www.stats.gla.ac.uk/~raya/howscalefree/howscalefree.html. Alternatively, a R-web interface is available on exgen.ma.umist.ac.uk/ under Public Tools.

3. Results

We have examined more than ten published datasets with biological interactions. The datasets include six protein-protein interaction networks data (Uetz et al. 2000; Schwikowski et al. 2000; Ito et al. 2001; Li et al. 2004; Rain et al. 2001; Giot et al. 2003), gene interaction networks obtained experimentally by genome wide location analysis (Lee et al. 2002) and synthetic lethal interactions (Tong et al. 2004) and networks constructed from databases (Guelzim et al. 2002) as well as metabolic interaction networks (Ma and Zeng, 2003; see Supplementary Table1). We have also analyzed gene expression networks constructed from the microarray data (combined yeast cell cycle data by Spellman et al. 1998) by thresholding Pearson correlations of pairs of genes, similarly to (Bergmann et al. 2004; van Noort et al. 2004). Most of these networks have been reported to be scale-free by fitting a straight line on a log-log plot (Figure 2).

Datasets	$\hat{\gamma}$	P -value
Uetz	2.05	0.0004
Schwikowski	1.87	$<10^{-4}$
Ito	2.02	$<10^{-4}$
Li	2.31	0.0014
Rain	1.61	$<10^{-4}$
Giot	1.53	$<10^{-4}$
Tong	1.45	$<10^{-4}$
Lee	2.03	$<10^{-4}$
Guelzim	1.56	$<10^{-4}$
Spellman/Cho	1.10(0.96)	$<10^{-4}$ ($<10^{-4}$)

Table 1: Exponents of the power-law and scale-free p-values for published protein and gene interaction datasets. This table contains results of our re-analysis of 10 published datasets that are referred here by the first author of the paper. These datasets include several experimentally obtained protein-protein interaction data (Uetz et al. 2000; Schwikowski et al. 2000; Ito et al. 2001; Li et al. 2004; Rain et al. 2001; Giot et al. 2003) for different organisms, gene interactions measured experimentally (Lee et al. 2002; Tong et al. 2004), constructed from the database (Guelzim et al. 2002) and constructed by thresholding pair-wise correlations with threshold equal 0.7 and 0.6 (values in parentheses) from the combined yeast cell cycle dataset (Spellman et al. 1998). For each dataset the power-exponent γ is calculated by maximum likelihood. The P -value summarizes the evidence against the Null-hypothesis that the network is scale free: these tiny ($<10^{-6}$) p-values suggest that the network is not consistent with a scale-free network.

However, the statistical analysis reported in Table 1 found that all these interaction dataset networks differ significantly from the power-law distribution based on chi-squared goodness-of-fit test (Analysis). These results clearly demonstrate that these interaction datasets do not follow the power-law distribution, in contrast of what has been commonly assumed or estimated from the log-log fit.

It has been sporadically reported (Jeong et al. 2001; Giot et al. 2003) that the connectivity distribution in some biological networks might be better described by a *truncated power-law*, i.e. a power-law regime followed by a sharp drop-off, $p(k) \sim k^{-\gamma} \exp(-k/k_c)$, $k \geq 1$. The parameter k_c is the cut-off, such that the number of connections k is less than expected for pure scale-free networks for $k > k_c$ and the behaviour of the network is approximately scale-free within the range $1 \leq k < k_c$. By fitting the truncated power-law to the interaction datasets (see Analysis), we have found that several datasets are consistent with the truncated power-law (Table 2; P -value > 0.1).

Datasets	$\hat{\gamma}$	\hat{k}_c	P -value
----------	----------------	-------------	------------

Uetz	1.60	8.7	0.370
Schwikowski	1.26	6.2	0.105
Ito	1.79	26	$<10^{-4}$
Li	2.10	19.5	0.018
Rain	1.12	11.5	0.200
Giot	1.09	20	0.001
Tong	0.96	23.7	$<10^{-4}$
Lee	1.96	294	$<10^{-4}$
Guelzim	1.18	15	$<10^{-4}$
Spellman/Cho	1.07(0.78)	73(99)	0.700 (0.100)

Table 2: Parameters of the truncated power-law and truncated power-law p-values for published protein and gene interaction datasets. The parameters of the truncated power-law distribution, γ and k_c , are calculated by maximum likelihood method for all 10 interactions datasets. Here the P -value summarizes the evidence against the truncated power-law distribution.

However, among those that can indeed be drawn from the truncated power-law distribution, the estimates of the cut-off parameter k_c have small values: this means that these networks exhibit scale-free behaviour only in a very small range of connectivities: $1 \leq k < k_c$. One exception is the gene co-expression network that has been constructed from the yeast microarray data (Spellman et al. 1998) by thresholding pair-wise correlations between gene profiles (last column in Table 2). Unlike other networks in this study, the yeast co-expression network is not a network of physical interactions. This network connects genes that might participate in the same biological processes. A large proportion of false positives in this network might contribute to its overall topology. In any case, this co-expression network does not belong to the class of scale-free networks but it might belong to a class of networks described by a truncated power-law.

Scale-free networks and broad-scale networks (characterized by truncated power-law) belong both to the class of small-world networks, but are of different types. Amaral et al. (2000) claim that such different types of networks are likely to be generated by different mechanisms: a plausible mechanism to explain the difference between two classes might be the constraints of node ageing and cost of adding new links. In terms of biological networks, their topology should give clues for the evolutionary mechanisms that have created those networks. Networks described by the truncated power-law can be generated by mechanisms that include evolutionary drift (Przytycka and Yu, 2004), while scale-free networks are consistent with duplication events (Rzhetsky and Gomez, 2001; van Noort et al, 2004).

4. Conclusions

The idea of scale-free behaviour and self-similarity at different levels of the network combines both the simplicity of the underlying principle and the complexity of the actual architecture of the network. It is therefore a very appealing concept, one that suggests a common underlying principle for different types of networks.

In this paper, however, we have found that the number of connections in biological networks significantly differs from the power-law distribution and these networks are not scale-free. Our conclusions are in resonance with those of Stumpf et al. (2005a) who have recently demonstrated by sampling properties of network degree distributions that the inferences about the scale-free nature of biological networks may have to be treated with some caution. In addition, Stumpf et al. (2005b,c) applied formal statistical model selection methods to determine with functional form best describes degree distributions of protein and metabolic networks from different organisms. These authors showed that that simple scale-free models do not provide an adequate description of real network data.

Therefore, the evolutionary mechanisms, such as gene duplication, evolution of new gene functions and gene loss, suggested to generate scale-free networks (Rzhetsky and Gomez, 2001; van Noort et al. 2004) do not by themselves adequately describe the appearance of biological networks. If the distributions are not scale-free, additional hypotheses are required to explain the underlying probability distributions (Kuznetsov et al. 2002). Recently, it has been demonstrated that divergent evolutionary drift, which is a plausible evolutionary mechanism, is not compatible with scale-free models (Przytycka and Yu, 2004).

Some of the conclusions drawn from supposedly scale-free behaviour are potentially misleading. Uncritically assuming that a network is scale-free might therefore lead to mistaken conclusions, for example that “the yeast co-expression network can be explained by a simple model” (van Noort et al. 2004). Also, randomly selected subsets of edges from a network that is not scale-free will not generally follow a power-law distribution and the interaction space as a whole will not have the same distribution shape as any major subset, as has been used for considerations of optimal design of protein-protein interaction experiments (Lappe and Holm, 2004). Moreover, the properties of a known part of a network that is not scale-free cannot be used directly to assess the properties of an unknown part of the network, as was claimed in Guelzim et al (2002) and Wolf et al (2002).

It is worth pointing out that the main qualitative results, i.e. small-world and lethality & centrality properties, still hold true independently of the precise mathematical form of the connectivity law and of their lack of scale-free behaviour for all of the datasets mentioned in this paper. Examples of other distributions with similar qualitative properties (i.e. existence of a few hubs and many nodes with a few connections) include truncated power-law (Jeong et al. 2001; Giot et al. 2003), generalized Pareto law (Kuznetsov et al. 2002; Przytycka and Yu, 2004), stretched

exponential distribution (Laherrere and Sornette, 1999), geometric random graph (Przulj et al. 2004), geometric distribution or combinations of the above. None of the above distributions, however, yields scale-free behaviour for the entire network. It is therefore crucial to generate a richer set of evolutionary models of biological networks, which are able to model and validate the network properties separately.

The relation between biology and network topology is very much an open question. At this stage, it is vitally important to get the topology right in order to make any future conclusions about the actual biology accessible and to start exploring complex issues of evolution and design principles of biological networks.

References

1. Alm E. and Arkin A. Biological networks. *Current Opinion in Structural Biology*, 13:193-202 (2003).
2. Aloy P. and Russell RB. Taking the mystery out of biological networks. *EMBO reports*, **5**(4), 349-350 (2004).
3. Agrawal H. Extreme self-organization in networks constructed from gene expression data. *Phys Rev Lett.*, **89**, 268702 (2002).
4. Albert R, Jeong H, and Barabasi AL. Error and attack tolerance of complex networks. *Nature*. 406:378-82 (2000).
5. Amaral L.A., Scala A., Barthelemy M. and Stanley H.E., Classes of small-world networks. *Proc Natl Acad Sci U S A.*, **97**,121,11149-52 (2000).
6. Barabasi A.L. & Oltvai Z.N. Network biology: understanding the cell's functional organization, *Nature Review Genetics*, **5**,2101-2113 (2004).

7. Bergmann S., Ihmels J. and Barkai N. Similarities and differences in genome-wide expression data of six organisms, *PLoS Biol*, **2**, 1 (2004).
8. Dokholyan NV, Shakhnovich B, and Shakhnovich EI. Expanding protein universe and its origin from the biological Big Bang. *Proc Natl Acad Sci U S A*. **99(22)**:14132-14136. (2002) .
9. Featherstone D.E. and Broadie K. Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network, *Bioessays*, **24**,3,267-274, (2002).
10. Fell D.A. and Wagner A. The small world of metabolism. *Nature Biotechnology*. **18**, 1121-1122 (2000).
11. Guelzim N., Bottani S., Bourguin P. and Kepes F. Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genetics*, **31**, 60-63 (2002).
12. Hoyle D.C., Rattray M., Jupp R. and Brass A. Making sense of microarray data distributions. *Bioinformatics* **18**, 576-584 (2002).
13. Huang S. Back to the biology in system biology: what can we learn from biomolecular networks. *Brief. Funct. Genomic Proteomic*. **2(4)**:279-297 (2004).
14. Ito T., Chiba T, Ozawa R, Yoshida M, Hattori M and Sakaki Y. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA*, **98**, 4569-4574 (2001).
15. Jeong H., Tombor B., Albert B., Oltvai Z.N. and Barabasi A.L. The large-scale organization of metabolic networks, *Nature*, **407**, 6804,651-654 (2000).

16. Jeong H., Mason S.P., Barabasi A.L. and Oltvai Z.N. Lethality and centrality in protein networks, *Nature*, **411**,683341-42 (2001).
17. Kuznetsov V.A., Knott G.D. and Bonner R.F. General statistics of stochastic process of gene expression in eukaryotic cells. *Genetics*, **161**, 1321-1332, (2002).
18. Lappe M. and Holm L. Unraveling protein interaction networks with near-optimal efficiency. *Nature Biotech.* **22**, 98-103 (2004).
19. Laherrere J. and Sornette D. Stretched exponential distributions in nature and economy: fat tails with characteristic scales. *European Physical Journal B*, **2**, 525-539 (1999).
20. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.R., Thompson, C.M., Simon I., Zeitlinger J., Jennings, E.G., Murray, H.L., Gordon, D.B., Ren, B., Wyrick, J.J., Tagne, J., Volkert T.L., Fraenkel, E., Gifford D.K., and Young, R.A. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*. **298**, 799-804 (2002).
21. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T, Goldberg DS, Li N, Martinez M, Rual JF, Lamesch P, Xu L, Tewari M, Wong SL, Zhang LV, Berriz GF, Jacotot L, Vaglio P, Reboul J, Hirozane-Kishikawa T, Li Q, Gabel HW, Elewa A, Baumgartner B, Rose DJ, Yu H, Bosak S, Sequerra R, Fraser A, Mango SE, Saxton WM, Strome S, Van Den Heuvel S, Piano F, Vandenhoute J, Sardet C, Gerstein M, Doucette-Stamm L, Gunsalus KC, Harper JW, Cusick ME, Roth FP, Hill DE, Vidal M. A map of the

- interactome network of the metazoan *C.elegans*. *Science*. **303**, 540-543, 2004.
22. Luscombe N.M., Qian J., Zhang Z. Johnson T. and Gertein M. The dominance of the population by a selected few: power-law behaviour applies to a die variety of genomic properties. *Genome Biology*. 3, 1-7 (2002). [16]
 23. Newman M.E.J. The Structure and Function of Complex Networks, *SIAM Review* 45, 167-256 (2003).
 24. Ma H.W. and Zeng A.P. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics*, **19**,1423-1430 (2003).
 25. Maslov S. and Sneppen K. Specificity and stability in topology of protein networks. *Science*, **296**, 910-913 (2002).
 26. Przulj N.C, Corneil D.G. and Jurisic I. Modeling interactome: scale-free or geometric. *Bioinformatics*, **20**, 3508-3515 (2004).
 27. Przytycka T. and Yu Y-K. Scale-free networks versus evolutionary drift. *Comput Biol Chem.*, **28**, 257-64 (2004).
 28. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V, Chemama Y, Labigne A, Legrain P. The protein-protein interaction map of *Helicobacter pylori*. *Nature*, **409**, 211-215 (2001).
 29. Rzhetsky A. and Gomez S.M. Birth of scale-free molecular networks and of the distinct DNA and protein domains per genome, *Bioinformatics*,**17**,988-986B (2001).

30. Schwikowski B., Uetz P. and Fields S. A network of protein-protein interactions in yeast, *Nature Biotechnol.*, **18**,1257-1261 (2000).
31. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell*, **9**, 3273-3297 (1998).
32. Stumpf MPH, Wiuf C, May RM. Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proc Natl Acad Sci U S A.*, **102**, 4221-4224 (2005a).
33. Stumpf and Ingram PJ. Probability models for degree distributions of protein interaction networks. *Europhys. Lett.*, **71** (1), pp. 152-158 (2005b)
34. Stumpf MPH, Ingram PJ, Nouvel I, Wiuf C. Statistical model selection methods applied to biological networks. *Transactions in Computational Systems Biology*, in press (2005c).
35. Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, Chen Y, Cheng X, Chua G, Friesen H, Goldberg DS, Haynes J, Humphries C, He G, Hussein S, Ke L, Krogan N, Li Z, Levinson JN, Lu H, Menard P, Munyana C, Parsons AB, Ryan O, Tonikian R, Roberts T, Sdicu AM, Shapiro J, Sheikh B, Suter B, Wong SL, Zhang LV, Zhu H, Burd CG, Munro S, Sander C, Rine J, Greenblatt J, Peter M, Bretscher A, Bell G, Roth FP, Brown GW, Andrews B, Bussey H, Boone C. Global mapping of the yeast genetic interaction networks. *Science*. **303**,565,808-813 (2004).

36. Ueda HR, Hayashi S, Matsuyama S, Yomo T, Hashimoto S, Kay SA, Hogenesch JB, Iino M. Universality and flexibility in gene expression from bacteria to human. *Proc Natl Acad Sci U S A.* **101**, 3765-9 (2004).
37. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*, *Nature*, **403**,6770,623-627 (2000).
38. van Noort V., Snel B. and Huynen M.A. The yeast coexpression network has a small-world, scale-free architecture and can be explained by a simple model. *EMBO reports*, **5**, 280-284 (2004).
39. Wuchty S. Scale-free behavior in protein domain networks. *Mol.Bio.Evol.* **18**, 1694-1702 (2001).

Appendix

In order to compute the power-law exponent and to test the hypothesis on whether the network is scale-free, we assumed that the connectivities of all the nodes in a large network are independent. To evaluate whether this assumption has an effect on our conclusions, we perform an alternative analysis, which relaxes this independence assumption. We note that connectivities of nodes in a sub-network are less dependent than the connectivities in the full network.

For each sub-network of randomly picked nodes constructed from the dataset network, the power-law exponent estimate $\hat{\gamma}$ and the usual chi-square test statistics $T = \sum_{k=1}^{k^*} (O_k - E_k)^2 / E_k$ were calculated as described in the Analysis section. Each sub-network contains one-tenth of the nodes in the whole network. Such sub-networks were sampled from the dataset network 1000 times, yielding the test statistic and the estimate for $\hat{\gamma}$, averaged over all sub-networks.

Then, 1000 networks were simulated with the same number of nodes as in the dataset network and the connectivities following the power-law with the exponent $\hat{\gamma}$. The procedure of sampling subnetworks from each of the simulated networks was repeated 1000 times, resulting in 1000 values of the chi-squared statistic for each simulated network. A one-side p-value for the dataset network was then computed.

The results for all the datasets studied in the paper are presented in the Table 3 supporting the conclusion that there is a lot of evidence that the experimental networks are not scale-free.

Datasets	$\hat{\gamma}$	P -value
Uetz	1.96	$<10^{-3}$
Schwikowski	1.79	$<10^{-3}$
Ito	1.98	$<10^{-3}$
Li	2.26	$<10^{-3}$
Rain	1.52	$<10^{-3}$
Giot	1.48	$<10^{-3}$
Tong	1.39	$<10^{-3}$
Lee	2.03	0.007
Guelzim	1.52	$<10^{-3}$
Spellman/Cho	1.08(0.94)	$<10^{-3}$ ($<10^{-3}$)

Table 3: Exponents of the power-law and scale-free p-values for published protein and gene interaction datasets, calculated under relaxing the independence assumptions.