Eur. Phys. J. B **57**, 75–80 (2007)
DOI: 10.1140/epjb/e2007-00145-0

THE EUROPEAN PHYSICAL JOURNAL B

A universal power law and proportionate change process characterize the evolution of metabolic networks

S. Singh[1], A. Samal[1], V. Giri[1], S. Krishna[2,a], N. Raghuram[3], and S. Jain[1,4,5,b]

[1] Department of Physics and Astrophysics, University of Delhi, Delhi 110007, India
[2] National Centre for Biological Sciences, UAS-GKVK Campus, Bangalore 560065, India
[3] School of Biotechnology, GGS Indraprastha University, Delhi 110006, India
[4] Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560064, India
[5] Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Received 5 December 2006
Published online 23 May 2007 –

Abstract. Biological and social systems have been found to possess a non-trivial underlying network structure of interacting components. An important current question concerns the nature of the evolutionary processes that have led to the observed structural patterns dynamically. By comparing the metabolic networks of evolutionarily closeby as well distant species, we present results on the evolution of these networks over short as well as long time scales. We observe that the amount of change in the reaction set of a metabolite across different species is proportional to the degree of the metabolite, thus providing empirical evidence for a 'proportionate change' mechanism. We find that this evolutionary process is characterized by a power law with a universal exponent that is independent of the pair of species compared.

PACS. 89.75.Hc Networks and genealogical trees – 87.23.Kg Dynamics of evolution – 89.75.Da Systems obeying scaling laws

1 Introduction

Several new structural patterns have been discovered in diverse biological, social and information networks, but the evolutionary dynamics that lead to such structures are still poorly understood. For example, metabolic networks, the best studied large scale networks in biology, are known to have a power law degree distribution [1,2], and the exponent γ is observed to be the same for all species (for a review see [3]). However, empirical evidence elucidating the nature of the process that gives rise to such structure is lacking. In this paper we present empirical facts about evolution based on a comparative study of metabolic networks of various organisms. In particular we report that the evolutionary process is itself characterized by a universal power law.

Various hypotheses regarding evolutionary mechanisms have been proposed and explored in mathematical models of evolving networks. For growing networks, a 'preferential attachment' of new nodes to higher degree nodes [4] as well as a 'proportionate change' mechanism [5,6] whereby nodes with higher degree experience proportionately higher changes in degree has been proposed to account for the power law degree distribution of the network. The latter process can lead to robust exponents [6]. However, it is not clear whether this hypothesis is applicable to the evolution of metabolic networks. For one, the metabolic network is not a growing network; during the course of evolution the number of metabolites has remained in the range of a few hundred to about a thousand for all organisms [7,8]. Furthermore, so far no concrete evidence has been presented for a preferential attachment or proportionate change process during its evolution. Our work also presents empirical evidence for a proportionate change process in metabolic network evolution.

2 Metabolic networks

We downloaded a database of metabolic networks of 107 organisms [8]. This contains organisms from all three domains: eukaryotes, prokaryotes and archaea, arranged in 15 groups including animals, plants, fungi, proteobacteria, firmicutes, and others. Organisms in different groups

[a] *Present address:* Niels Bohr Institute for Astronomy, Physics and Geophysics, Blegdamsvej 17, Copenhagen 2100, Denmark.
[b] e-mail: jain@physics.du.ac.in

are evolutionarily distant, while those in the same group are relatively closeby. We selected one species from each group (typically the one having the largest number of metabolites) and compared the metabolic networks of all 15 species pairwise (105 pairs of distant species). We also compared specific pairs of nearby species (within the same group).

The metabolic network of a given species is the set of catalysed chemical reactions that can take place in the organism through which it converts 'food molecules' into certain other types of molecules needed by its cells. The above database contains a list of 5275 metabolic reactions, and for each reaction gives its participating metabolites, chemical equation, whether it is reversible or not, and whether it exists or not in each species.

3 A 'local' measure of distance between metabolic networks

For a given pair of species, say A and B, consider the set M of metabolites that participate in one or both metabolic networks. For every metabolite m in M, we now define Δk^m_{AB}, a measure of the distance between the two networks. Let R^m_A (R^m_B) be the set of reactions in which m participates in the metabolic network of A (B). The number k^m_A (k^m_B) of reactions in R^m_A (R^m_B) is the degree of m in the species A (B). Here we consider only the undirected degree of a metabolite, i.e., we do not distinguish whether the metabolite participates as a reactant or a product. Reversible reactions (forward and reverse pair) in which a metabolite participates are treated as a single reaction for calculating its degree. If m occurs in only one of the species (say, A) and not the other (B), then R^m_B is the null set and $k^m_B = 0$.

Consider the reactions in $R_A \cap R_B$. (For brevity we will often drop the superscript m as being understood.) The reactions in $R_A \cap R_B$ represent the links of the metabolite that are common to both species, and hence k^m_{AB}, the size of this set, is a measure of how much the reaction set of this metabolite has remained 'conserved' in the evolution leading to species A and B from their last common ancestor. (This is a kind of 'local measure' of conservation, from the vantage point of this metabolite inside the network.) Similarly, the set $(R_A \cup R_B)\backslash(R_A \cap R_B)$, that is, the set of reactions in $R_A \cup R_B$ that are not in $R_A \cap R_B$, or equivalently those reactions of this metabolite that are in one network but not the other, is a local measure of the divergence between the two networks. The size of the latter set will be referred to as the divergence of the reaction sets of this metabolite between species A and B, and will be denoted Δk^m_{AB}. Note that Δk^m_{AB} is different from the magnitude of $k^m_A - k^m_B$. For example R^m_A and R^m_B can be different sets of reactions with the same number of reactions in which case $k^m_A - k^m_B = 0$ while $\Delta k^m_{AB} \neq 0$. Δk^m_{AB} is a local measure of the difference between the two networks that takes into account the *identity* of reactions and not just their number.

4 Metabolic network evolution is characterized by a universal power law

4.1 Comparison of distant species: Evolution on long time scales

We computed the degree distribution $P(k)$ for each of the 15 organisms as well as the 'divergence probability distribution' $Q(\Delta k)$ for each of the 105 pairs. By definition, for any pair (A, B), $Q_{AB}(\Delta k) \equiv n_{AB}(\Delta k)/|M|$, where $|M|$ is the number of metabolites in M and $n_{AB}(\Delta k)$ is the number of metabolites in M for which $\Delta k^m_{AB} = \Delta k$. The cumulative divergence distribution for pairs of distant organisms is shown in Figure 1 and compared with the cumulative degree distribution. The figure shows that $Q(\Delta k) \sim (\Delta k)^{-\gamma'}$ with $\gamma' = \gamma$ up to statistical uncertainties in both the exponents. That the degree distribution of two species follows the power law $P(k) \sim k^{-\gamma}$ is a statement of the present structure of the two metabolic networks. This in no way implies that $Q(\Delta k)$ should also follow a power law with the same exponent. The latter is a distinct statement about the dynamical process that leads to the present structure. That the $Q(\Delta k)$ distribution has the same form for all 105 pairs of distant species considered reflects a universal property of the evolutionary process.

4.2 Evolution on shorter time scales

A comparison of distant species reveals features of the evolutionary process over long time scales. In order to study the process over short time scales we compared nearby species (that were in the same group). The result of three such comparisons is shown in Figure 2. As expected, for each pair of nearby species the absolute divergence is smaller than for distant species. This is evident from the fact that the $Q(\Delta k)$ curves are well below the $P(k)$ curves in Figure 2, in contrast to Figure 1 where they are much closer, and that larger values of Δk are absent in Figure 2. However, it can be seen that the $Q(\Delta k)$ still follows a power law with almost the same exponent. This suggests that this feature of the evolutionary process is also valid over short evolutionary time scales.

5 Metabolic network evolution follows a 'proportionate change' process

We explored the relationship between Δk for a metabolite across a pair of species, and its degree in each of those species. In particular, one can ask for the conditional probability $P(\Delta k|k)$ for a metabolite to have a reaction set divergence Δk across a pair of species, given that its average degree in the two species is k. We found a positive and approximately linear correlation between Δk and the degree of a metabolite (Figs. 3 and 4). Thus the difference in the reaction set of a metabolite across species is proportional to the size of the set.

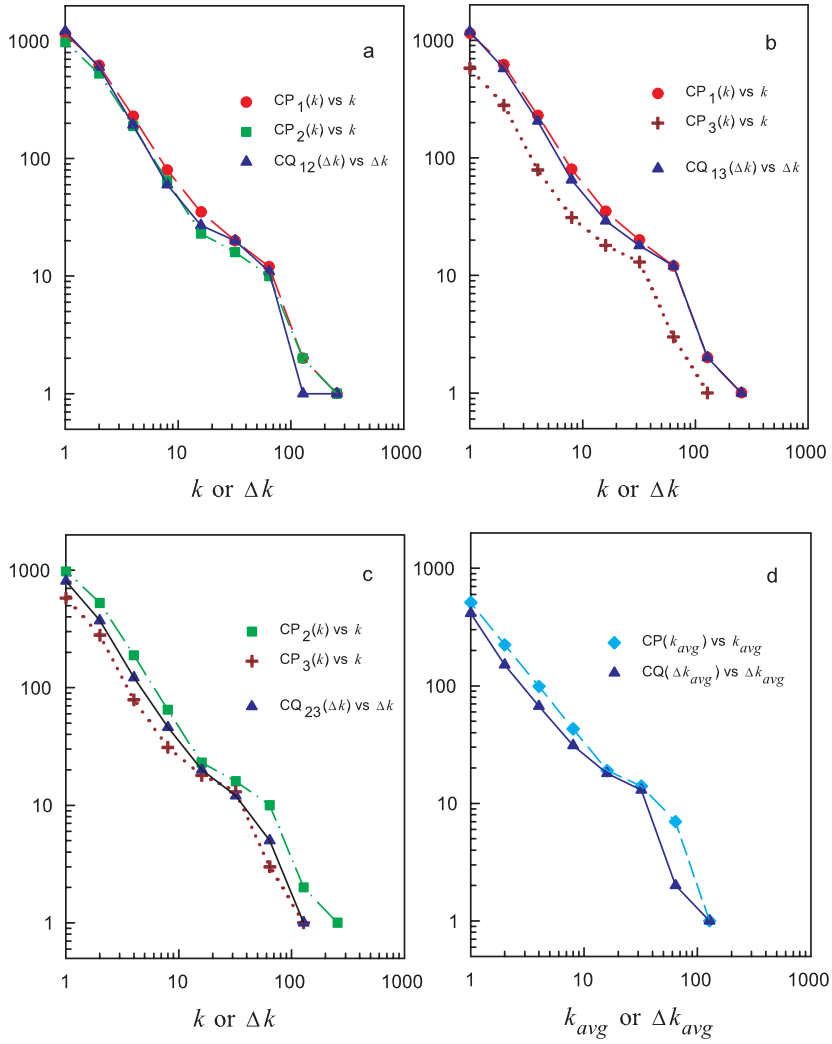

Fig. 1. The cumulative distribution of the divergence of reaction sets, $CQ(\Delta k)$, defined below, for evolutionarily distant organisms. This is compared with the cumulative degree distribution $CP(k)$ of those organisms. (a–c) Exhibit the comparison for three pairs of species. The species are the eukaryote *Homo sapiens*, the prokaryote *Escherichia coli* and the archaean *Methanosarcina mazei* indexed as 1, 2, 3 respectively. (d) The cumulative distribution of the degree of a metabolite averaged over 15 species (k_{avg}) and distribution of the divergence of the reaction sets of a metabolite averaged across the 105 pairs of species (Δk_{avg}). Points on the $CP_A(k)$ vs k curves ($CQ_{AB}(\Delta k)$ vs. Δk curves) represent the number of metabolites with k_A^m (Δk_{AB}^m) greater than or equal to 2^{r-1}, $r = 1, 2, \ldots$ The value of the exponent γ' $(= 1 + |\text{slope}|)$ obtained from the least square fit value of the slope of the respective curve $\pm$ standard error (one standard deviation) arising from the scatter of the points plotted in the figure is (a) 2.31 ± 0.09 ($Q_{12}(\Delta k)$); (b) 2.26 ± 0.06 ($Q_{13}(\Delta k)$); (c) 2.31 ± 0.06 ($Q_{23}(\Delta k)$); (d) 2.19 ± 0.08 ($Q(\Delta k_{avg})$). The value of the exponent in $P(k) \sim k^{-\gamma}$ is $\gamma = 2.27 \pm 0.05$ ($P_1(k)$), 2.24 ± 0.05 ($P_2(k)$), 2.26 ± 0.07 ($P_3(k)$), and 2.26 ± 0.08 ($P(k_{avg})$). Across the 15 organisms considered, γ ranges from 2.22 to 2.32 with a mean of 2.27, while across the 105 pairs γ' ranges from 2.23 to 2.46 with a mean of 2.31.

A random 'proportionate change' type process has been proposed in models of growing networks [5,6]. We propose the following general definition of a proportionate change type process valid for random/non-random evolution in growing/non-growing networks and for models as well as real networks: In an evolving network existing nodes can lose some of their existing links or gain new links. If in a certain time interval of evolution, the number of links lost plus gained by nodes is typically in proportion to the links they had at the beginning of the interval, the network evolution will be said to occur via a 'proportionate change' type process in that time interval. It does not matter, for purposes of this definition, what the underlying process causing the change is. It could be a random process or a highly designed process. As long as the net change in the links of nodes ends up being linearly correlated with their initial degree, the evolutionary process will be referred to as a 'proportionate change' type process. In this sense, our result in the previous paragraph is evidence of a 'proportionate change' type process in the evolution of metabolic networks. We remark that network evolution does not have to follow such a scheme; for

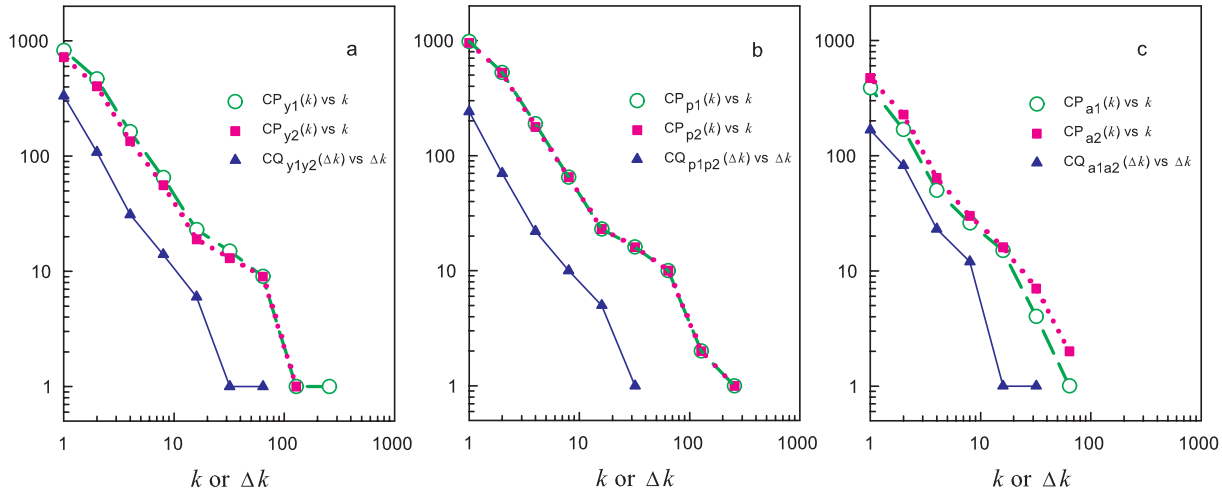

Fig. 2. $CQ(\Delta k)$ and $CP(k)$ compared for evolutionarily closeby species. (a) Compares two yeasts, *Saccharomyces cerevisiae* (labeled as y1; $\gamma = 2.26 \pm 0.07$), and *Schizosaccharomyces pombe* (y2; $\gamma = 2.26 \pm 0.09$), and γ' is found to be 2.46 ± 0.10 ($Q_{y1y2}(\Delta k)$). (b) Compares two proteobacteria, *E. coli* (p1; $\gamma = 2.24 \pm 0.05$) and *Salmonella typhimurium* (p2; $\gamma = 2.23 \pm 0.05$); $\gamma' = 2.49 \pm 0.09$. (c) Compares two archaea, *Pyrococcus horikoshi* (a1; $\gamma = 2.37 \pm 0.08$) and *Pyrococcus furiosus* (a2; $\gamma = 2.27 \pm 0.05$); $\gamma' = 2.63 \pm 0.18$.

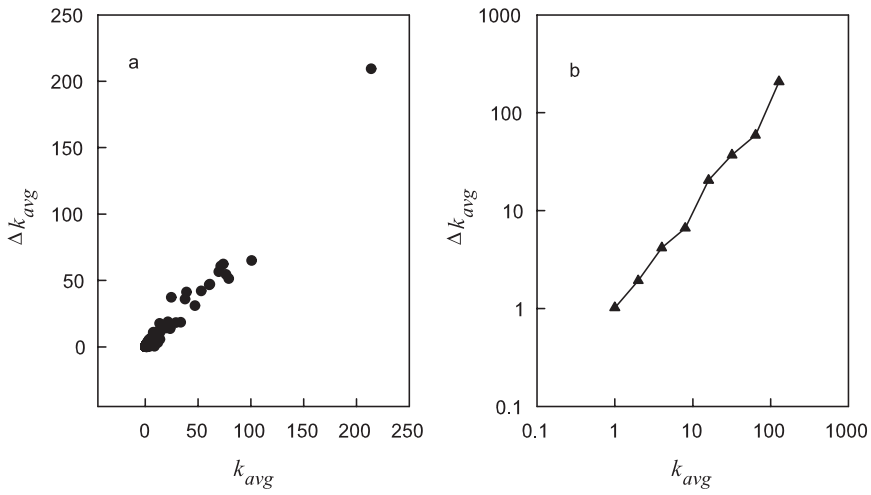

Fig. 3. Positive and approximately linear correlation between Δk and k (a) Scatter plot (on a linear scale) of the average Δk of a metabolite across the 105 pairs of species versus its average degree across the 15 (distant) species. The lone point on the extreme right is a single highly connected metabolite, the hydrogen ion. (b) The same on a logarithmic scale where metabolites are placed in logarithmic bins according to their average degree, and the average Δk for a bin is computed by averaging over all 105 pairs of organisms for a given metabolite and then averaging over all metabolites in the bin. The slope of the least square fitted straight line $\pm$ the standard error of the deviation of points in the figure from the fit is 1.08 ± 0.03.

example a simple evolution rule that deletes randomly chosen nodes and/or adds new nodes that connect to randomly chosen existing nodes will not correspond to proportionate change. Thus our observation above captures a definite pattern in metabolic network evolution.

The result also provides insight into why the exponents γ' and γ might be equal or very close. For, let us assume for the moment a perfect correlation between Δk and k, i.e., $P(\Delta k|k) \sim \delta(\Delta k - f(k))$, or, equivalently, that $\Delta k = f(k)$ for some fixed one-to-one function f, and also that $P(k) \sim k^{-\gamma}$. Then the statement $f(k) \sim k^{\alpha}$ is equivalent to the statement $Q(\Delta k) \sim (\Delta k)^{-\gamma'}$, with $\gamma' = \gamma/\alpha$. In particular $\alpha = 1$ implies $\gamma' = \gamma$ and vice versa[1]. However this is not a complete explanation because, as is evident from Figures 3 and 4, there is stochasticity in the relation between Δk and k, and not perfect correlation.

[1] We remark that given the statistical uncertainty in the exponents, our results are consistent with a value of α slightly different from 1 and γ' slightly different from γ.

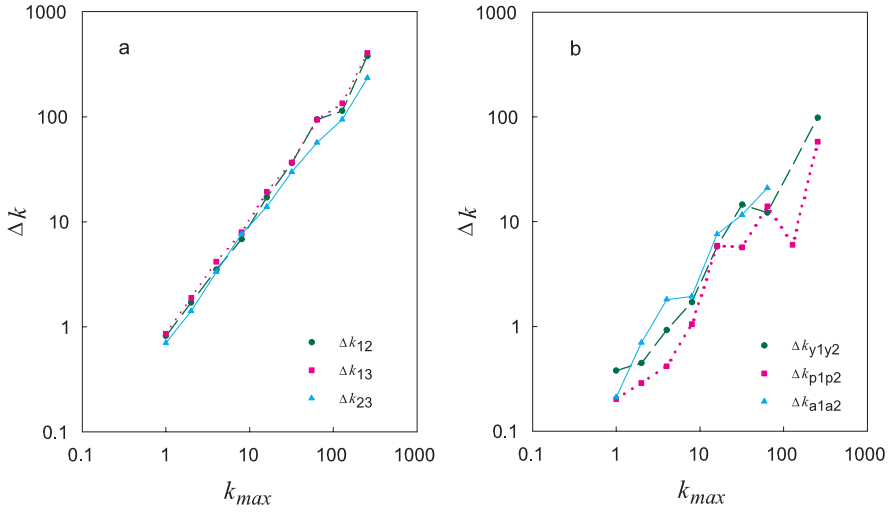

Fig. 4. Δk versus k of a metabolite for three pairs of (a) distant species and (b) close-by species. The species and their indices in (a) are the same as in Figures 1a–1c and in (b) are the same as in Figures 2a–2c. For each pair of species (A, B), the x-axis represents $k_{\max} = \max(k_A, k_B)$. The slopes of the three lines in (a) are 1.09 ± 0.03 (1, 2); 1.08 ± 0.02 (1, 3); and 1.03 ± 0.02 (2, 3) and in (b) the slopes of best fit lines are 1.03 ± 0.07 (y1, y2); 0.97 ± 0.12 (p1, p2); and 1.07 ± 0.08 (a1, a2).

6 Possible molecular mechanisms for proportionate change in metabolic network evolution

The linear correlation described above is an overall characterization of the evolutionary process. A deeper understanding would require going into the mechanisms by which such a correlation comes about, as well as into the departures from the statistical pattern. A metabolic reaction is catalyzed by an enzyme to which the reactant molecules bind at specific sites in a 3-dimensional geometry. Hence metabolic network evolution ultimately rests on mechanisms of enzyme structure evolution [9], which in turn involves the molecular evolution of genes that code for the enzymes. The latter is governed both by random processes as well as the forces of selection. The following random processes that are biologically plausible can in principle give rise to a proportionate change in metabolic networks: a metabolite with high degree binds to several enzymes that catalyze its reactions; if a gene corresponding to one of these enzymes mutates in a manner that disturbs the binding site of this metabolite on the enzyme, the corresponding reaction could be lost. The more enzymes the metabolite binds to, the proportionately higher is the probability of losing its reactions through random mutations. On the other hand if the gene duplicates and diverges, that can introduce a new enzyme to which the metabolite binds and hence a new reaction for it to participate in. Large degree metabolites have a larger pool of interacting enzymes whose genes can duplicate, and hence if genes duplicate randomly, the number of new reactions a given metabolite participates in is also expected to be positively correlated with its degree. For two species A and B that have descended from a common ancestor, these processes would imply that $\Delta k \sim k$. Thus the same mechanisms, namely gene mutations and duplication-divergence, that have been considered as mechanisms for proportionate change and preferential attachment in protein interaction networks [10–16], could operate for metabolic networks also.

7 Discussion

The above mentioned mechanisms are attractive for their economy in explaining proportionate change: they invoke only random processes and do not invoke selection. However, selection also certainly shapes metabolism. There is evidence [17,18] that the amount of conservation or divergence in the reaction sets of particular metabolites depends upon the role played by those metabolites in the network, and that cells can direct the generation of potentially favourable mutations with greater probability than random [19]. Delineating the respective contribution of random/directed processes in proportionate change as well as other aspects of metabolic network evolution is an important task for the future. Some insight may be provided by the design of metabolic networks and the observation known to biochemists and emphasized recently [20,21] that the role played by a metabolite is reasonably well correlated with its degree. Our methods allow a systematic investigation of the deviation of individual metabolites from the null hypothesis of proportionate change, which also impinges on this question. We add that the method used in this paper to compare metabolic networks could be useful for comparing other labeled bipartite graphs.

We thank IUCAA Reference Centre, University of Delhi for computational infrastructure. We also thank S.N. Bose National Centre for Basic Sciences, Kolkata, and Centre for High Energy Physics, IISc, Bangalore for infrastructure and hospitality where part of this work was done. S.S. and A.S. acknowledge Senior Research Fellowships from the University Grants Commission (UGC) and the Council of Scientific and Industrial Research, respectively. V.G. acknowledges a Junior Research Fellowship from UGC.

References

1. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabasi, Nature (London) **407**, 651 (2000)
2. A. Wagner, D.A. Fell, Proc. R. Soc. Lond. B **268**, 1803 (2001)
3. A.-L. Barabasi, Z.N. Oltvai, Nature Rev. Genetics **5**, 101 (2004)
4. A.-L. Barabasi, R. Albert, Science **286**, 509 (1999)
5. B.A. Huberman, L.A. Adamic, Nature (London) **401**, 131 (1999)
6. B. Kahng, Y. Park, H. Jeong, Phys. Rev. E **66**, 046107 (2002)
7. M. Kanehisa, S. Goto, Nuc. Acids Res. **28**, 27 (2000)
8. H. Ma, A.-P. Zeng, Bioinformatics **19**, 270 (2003); `http://genome.gbf.de/bioinformatics/publications.html`
9. S. Schmidt, S. Sunyaev, P. Bork, T. Dandekar, Trends in Biochemical Sciences **28**, 336 (2003)
10. R.V. Sole, R. Pastor-Satorras, E. Smith, T.B. Kepler, Adv. Complex Syst. **5**, 43 (2002)
11. A. Vazquez, A. Flammini, A. Maritan, A. Vespignani, Com. Plex. Us. **1**, 38 (2003)
12. H. Jeong, Z. Neda, A.-L. Barabasi, Europhys. Lett. **61**, 567 (2003)
13. J. Kim, P.L. Krapivsky, B. Kahng, S. Redner, Phys. Rev. E **66**, 055101 (2002)
14. F. Chung, L. Lu, T. Gregory-Dewey, D.J. Galas, J. Comput. Biol. **10**, 677 (2003)
15. E. Eisenberg, E.Y. Levanon, Phys. Rev. Lett. **91**, 13870 (2003)1
16. S. Wuchty, Z.N. Oltvai, A.-L. Barabasi, Nature Genetics **35**, 176 (2003)
17. J.M. Cork, M.D. Purugganan, BioEssays **26**, 479 (2004)
18. R. Guimera, L.A.N. Amaral, Nature (London) **433**, 895 (2005)
19. L.H. Caporale, Annu. Rev. Microbiol. **57**, 467 (2003)
20. R. Tanaka, Phys. Rev. Lett. **94**, 168101 (2005)
21. L. Li, D. Anderson, R. Tanaka, J.C. Doyle, W. Willinger, e print `arXiv:cond-mat/0501169`