Update TRENDS in Biochemical Sciences Vol.29 No.7 July 2004

Research Focus

Strong links are important, but weak links stabilize them

Peter Csermely

Department of Medical Chemistry, Semmelweis University, H-1088 Budapest, Puskin str. 9, Hungary

Recently, a highly uneven distribution of metabolic network fluxes has been described in *Escherichia coli* that defines a high-flux backbone containing the most important (and most familiar) metabolic reactions. These findings might lead to a better understanding of the organization of the minimal gene set, and to a novel explanation for the allometric scaling laws. The scale-free distribution of the metabolic fluxes of *E. coli* highlights the importance of weak links in metabolic, and other, networks. Several proven examples show that weak links might have a general role in the stabilization of a variety of complex systems and, in addition, provide novel interpretations of the role of water in the potential energy landscape of the protein-folding process.

Network studies have received a lot of attention recently [1–4]. Whereas earlier work concentrated mostly on NETWORK TOPOLOGY (see Glossary) and established important rules on the DEGREE, MODULAR HIERARCHY and failures upon NETWORK DAMAGE, recent efforts have attempted elucidation of the dynamism of networks by examining the traffic along their connections. In an important step in these later studies, Barabasi and co-workers have reported a highly uneven distribution of METABOLIC NETWORK FLUXES in *Escherichia coli* [5]. By applying the method of FLUX-BALANCE ANALYSIS [6], which enables the estimation of the flux of each enzyme reaction, it was shown that *E. coli* metabolism is dominated by very few reactions with high fluxes. For example, the flux of succinyl coenzyme A synthetase – an important enzyme of the citrate cycle at the center of *E. coli* energy metabolism – accounted for almost 20% of all fluxes in *E. coli*, and was 10 000 times that of aspartate oxidase (an enzyme involved in amino acid metabolism). Barabasi and co-workers [5] determined that the distribution of fluxes followed a power law with the relative frequency of flux, ν, as a function of $(\nu + \nu_0)^{-\alpha}$, where ν_0 is an additive constant and the exponent, α, has a value of 1.5. Metabolic network fluxes can be regarded as strength indicators of the links between various elements of the metabolite (substrate or product) network of *E. coli*.

Key reactions in *E. coli* metabolism form a high-flux network

In-homogeneity of *E. coli* metabolic fluxes is global as well as local, which means that the greater the number of reactions relating to a given metabolite, the more likely it is that there is a key reaction among them that carries most of the flux [5]. These key reactions form a high-flux backbone of *E. coli* metabolism. The backbone resembles the traditional metabolism charts described in biochemistry textbooks. This is not a mere coincidence – the core of our biochemical knowledge has been gathered and built up by elucidating the most available reactions first, which are those found in the high-flux backbone. The in-homogeneity of fluxes was actually a fortune for early biochemical studies because, without them, the founding fathers of biochemical metabolism would have been lost for quite a while in the jungle of equal enzyme reactions. The differing importance of the cellular enzymes is probably related to the MINIMAL GENE SET concept of Mushegian and Koonin [7]. The minimal gene set refers to a set of essential genes that are needed for all organisms. In this concept, redundancy of genes is eliminated (in this context, 'genes' refers to a set of isoenzymes). Whether or not the high-flux reactions correspond to these essential genes and distinguish a subset of evolutionarily 'frozen' proteins will be an interesting subject of future studies.

The existence of the high-flux backbone also means that small fluxes are locally channeled to high-flux pathways. The essence of the flux-distribution and -connection scheme resembles that of several transportation networks such as rivers, trees and blood vessels. It will be of considerable interest to determine whether the observed flux distribution of an organism is related to the cellular version [8] of the surprisingly universal ALLOMETRIC SCALING LAWS (e.g. Kleiber's law), which show a $\frac{3}{4}$ power-law dependence of many biological properties on the body mass (i.e. $P = cM^{3/4}$, where P is the property, c is a constant and M is the body mass). A limited fraction of this empirical scaling law is often called the 'mouse-to-elephant curve', which describes the universality it represents by being equally valid to all mammals, from mice to elephants. Recently, however, the applicability of the formula has been extended further to cells and individual enzymes [8]. Although several approaches were used to derive these empirical scaling laws from the properties of transportation networks, the existence of a carefully orchestrated intracellular transportation network is not readily apparent. This shows the importance to attempt the 'channeling' of the power-law flux distribution into the existing explanations of scaling laws.

However, it must be noted that models always have limitations. Thus, the over-interpretation of the allometric scaling laws has been debated [9], and the approach of

Corresponding author: Peter Csermely (csermely@puskin.sote.hu).
Available online 10 June 2004

Glossary

Allometric scaling laws: A large variety of empirical scaling relationships that show self-similar behaviors of various complex systems, such as cells, organs and organisms, over a wide range of their mass. This relationship often follows a power law, $P = cM^{\alpha}$, where P is the property, c is a constant, M is the mass and α is a scaling exponent – usually 0.75. In Kleiber's law, P is the metabolic power, that is, the amount of energy per unit of time required by a living organism to remain alive.

Degree: The number of links of a given node in a network. The degree of distribution is an important property of a given network that shows whether the network is a random graph (with Gaussian degree distribution), a scaled graph (with exponential degree distribution), or whether the distribution of its degrees is a mix of both.

Flux-balance analysis: A method to calculate the metabolic status of a complex organism developed by Bernhard Palsson and co-workers for *Escherichia coli* [29]. Assuming that cellular metabolism is in a steady state and optimized for the maximal growth rate, flux-balance analysis enables the calculation the flux of each reaction using linear optimization.

Hit-and-run method: A method used by Almaas *et al.* [5] that randomly samples all possible solutions for the fluxes of chemical reactions under non-optimal conditions.

Metabolic network fluxes: The network representation of the metabolism of an organism (e.g. *Escherichia coli*) defines the various reaction substrates and products as elements (nodes) and the chemical reaction fluxes as links between them. The flux of a reaction (link) of this network represents the amount of substrate converted during a defined period of time. In the work by Almaas *et al.* [5], the flux vector was normalized to unity; therefore, fluxes were given as dimensionless units. Almaas *et al.* [5] define high flux as the flux that provides the largest incoming (or outgoing) flux contribution to each metabolite. This definition is justified by the observed highly uneven distribution of fluxes even in the local level.

Minimal gene set: The group of genes that have universal distribution and seem to be necessary for the current version of life on Earth.

Modular hierarchy: Many networks have modules (i.e. groups of nodes) that have a much higher number of connections between them than they have with the outside elements of the network. If the connections between the modules are uneven and one or more central module develops to which the other, peripheral modules are connected, we refer to it as modular hierarchy.

Network damage: An inhibition of network performance (e.g. efficient communication via the links of the network) after the removal or attenuation of nodes or links in the network. Scale-free networks are robust against random damage but are especially sensitive to the damage of their hubs, that is, nodes with many links. Cascading damage refers to damage that gradually propagates in the network.

Network topology: The topology of a network is the precise description of the links between the nodes (elements) of the network. Many special topologies are discriminated by their degree distribution.

Optimal stability: The optimal stability of a network can be conceived in two levels. (i) A network is stable if it shows a tendency to return to its original parameters after a perturbation. (This definition resembles to the Le Chatelier principle with the important difference that complex systems are almost never in a traditional equilibrium. Therefore, it is better to describe stability as robust behavior gravitating towards certain parameter sets, attractors of the network after perturbations.) (ii) At another level of stability, the network might leave the original attractors and shift to new ones, or even undergo a phase transition of its topology, for example, changing the scale-free degree distribution to a star or random degree distribution as the outside conditions become harder or easier, respectively. Here, the criterion for stability is to keep the network functional, that is, to preserve most of its connections. For cells or organisms, this means that the network stays alive. For social and ecosystem networks, this type of stability is often called resilience to discriminate it from the 'simpler', chemical type of stability.

Scale-free topology: A distribution of the degrees of network nodes that follows a power law (algebraic decay instead of, for example, exponential). These networks have many nodes with a low degree (i.e. few connected links); however, they also have a non-zero number of hubs (i.e. nodes with an unusually high number of links).

Self-organized criticality: An unstable state of networks in which tension develops as the network grows. The tension is released by an avalanche-type cascading change in the network when the system becomes critical (i.e. many of its elements behave identically as in phase-transitions). Both the extent and the probability of this critical behavior often follow a power law.

constructing metabolic networks using metabolites as nodes and reactions as links might give, to some degree, different data depending on whether the simplest inorganic or organic chemicals are included [10].

The scale-free distribution of *E. coli* metabolic fluxes is surprisingly robust

In the recent paper by Almaas *et al.* [5], the distribution of *E. coli* metabolic network fluxes described is surprisingly robust. Thousands of random choices from the nutrients of the *E. coli* menu did not change the power-law distribution $[(\nu + \nu_0)^{-\alpha}]$ under optimal growth conditions. Moreover, when 50 000 non-optimal states were examined only the exponent of the power-law flux distribution varied, the distribution itself remained stable. However, how well the applied HIT-AND-RUN METHOD [5] represents the sophisticated stress response of *E. coli* under changing conditions of sub-optimal growth remains an open question. Under stress, many of the regular enzyme reactions in *E. coli* are grossly diminished, which – in the first approximation – would imply an increase in the exponent from the optimal 1.5 because there is a shift of high-flux to low-flux reactions. Whether this shift actually occurs or whether fluxes of these transient, unstable regimes do not obey the power law distribution at all, deserves further investigation.

Interestingly, most of the adaptive processes of *E. coli* affect reactions of the high-flux backbone. Several pathways have a high flux under certain conditions but are turned off completely under different conditions. In addition, many reactions remain active under different conditions but show large flux-shifts. Almost all of these latter reactions belong to the high-flux backbone and have several stable flux values that are distinct from each other. However, despite these large changes in individual fluxes, the overall distribution of the *E. coli* flux network remains the same in all conditions examined. This suggests that the robustness of the SCALE-FREE TOPOLOGY distribution of the link number at a given node and the robustness of the link strength in the whole network are related [5,11]. Moreover, the scale-free distribution of the network representation in space (topology) and in energy (connection strength) can be extended to the time dimension as well because the distribution of stochastic events often follows a power law in complex networks, as demonstrated by *Per Bak* and co-workers showing the SELF-ORGANIZED CRITICALITY of these systems [12]. Whether the scale-free distributions of these three parameters are related to each other (self-organized criticality is just a consequence of the other two scale-free distributions) and whether they all reflect the self-organizing behavior of complex systems [13] remain exciting questions.

Weak links stabilize complex systems

The surprising resilience of the strength distribution of network connections revives another important lesson. If all the information regarding the existence of the high-flux backbone of *E. coli* and the relationship of these fluxes to the most important metabolic pathways were summarized, and the reader was asked which are the most important – strong links or weak links – the answer would be 'obviously strong!'. But this assumption is obviously wrong. Both types of links – strong and weak – are needed; in fact, a careful balance of both is required, which is reflected in the robustness of their strength distribution.

So it's clear: weak links are also important. Yet another question arises: it appears that the need for strong links is understood – without them there is no backbone, the network itself is not defined – but why are weak links required?

The answer comes from a variety of other studies. In essence: weak links are not as important as strong links in networks, but are crucial for their stabilization. Table 1 summarizes several networks in which weak links are crucial to the stabilization of complex systems. Most of these studies were done in social networks, where the stability of various systems, starting from companies and other smaller communities up to societies, was shown to depend on the number of informal, weak contacts between the participants of the networks. However, recently, other examples of the stabilizing role of weak links in animal communities and in complex ecosystems were also uncovered. Cells are well controlled thermodynamically and by their genetic background, however, weak links of, for example, their molecular chaperones [14] or p53 [15] might also be important for their stability.

Weak links also emerge with other network-stabilizing factors, such as modules [16] or degeneracy [17]. Modules are often connected by weak links, which can work as fuses and, by a temporary disconnection of modules, might prevent the propagation of cascading failures. However, degeneracy – the property of complex systems to perform the same tasks by different mechanisms – also involves the emergence of weak links. Degenerate network modules or motifs are different from each other, but have similar functions that are linked to the same set of target modules or motifs. Connection of dissimilar, degenerate modules to the same target modules requires weak links. Thus, degeneracy has a key role in network stability of a variety of networks [17].

Weak links seem to stabilize most networks, but they are not the only means by which a network can be stabilized. Networks might also change their topology, which can lead to a phase transition [18], that is, a change from a scale-free distribution to a star distribution (in which a central node attracts most of the links, like a dictator in centralized societies) or to a random graph, in which links are randomly formed between network elements (resembling to hunter-gatherer societies during periods when there is ample food supply). In cells, network topology is often altered by changes in gene expression.

The surprising variety of social and ecosystem networks that are stabilized by weak links, together with the emergence of weak links in general network-stabilizing scenarios (modules and degeneracy) and the conservation of the weak-link ratio obeying a power law at the cellular level [5], raises the possibility that weak links might be required to the stabilization of all complex systems. It seems that the first assumption should be turned around: weak links are the really important parts of networks. Wrong again. 'Over-stabilization' of networks leads to less noise within the network and to less diversity between various networks [19]. Less diversity of networks will decrease the amount of weak links between them. (Uniform networks can only make uniform links.) Extreme uniformity of constituent networks will destabilize the 'grand-network' that they form at a higher level of complexity. And so on. Networks balance on the brink of stability. Most networks exist owing to the OPTIMAL STABILITY of their constituent networks, and provide an important element of stabilization to the 'grand network' formed by them. Our optimal stability is an important part of (ultimately) the optimal stability of the universe. And we most probably need an optimal number of weak links in and around us to perform this evolutionary task.

The importance of a balanced distribution of weak and strong links can also be shown by the network representation of a potential energy landscape [20]. Here, nodes of the network represent local energy minima (e.g. those of protein conformations during the folding process), links between these energy minima correspond to the transition states between them. To define such a network representation of a potential energy landscape, multiple local energy minima are required because, without them, the elements of the network are lost. A link can be defined as strong if the transition state has a low activation energy and, conversely, a link is weak if the transition state has a high activation energy. In this context, weak links mean that the probability of a conformational change occurring between two corresponding nodes (e.g. protein conformations having a local energy minimum) is low because of the high activation energy between the two energy states. If we imagine the extreme situation of having only strong links between the local energy minima, all the activation energies will be low, and all the transitions will occur easily (Figure 1). This system would be undefined (unstable) because the protein would shift to the global energy minimum without any appreciable transition time in local energy minima (unrestricted, fast protein folding). By contrast, if there are only weak links between the energy minima, all local energy wells would behave as folding traps: the system becomes 'super stable', but the network itself is not defined because almost no folding pathways can be achieved in which the conformation gradually changes until reaching the global energy minimum. The folding pathway of larger proteins corresponds to the 'normal' network shown in Figure 1; the

Table 1. Examples of networks in which weak links are crucial to network stability

Weak links between	Stabilized network	Stabilized function	Refs
Animals	Animal community	Group survival	[23]
Citizens	Society	Social efficiency	[24,25]
Employees	Company	Firm efficiency	[26]
Consumers and preys	Ecosystem	Ecosystem stability	[27,28]

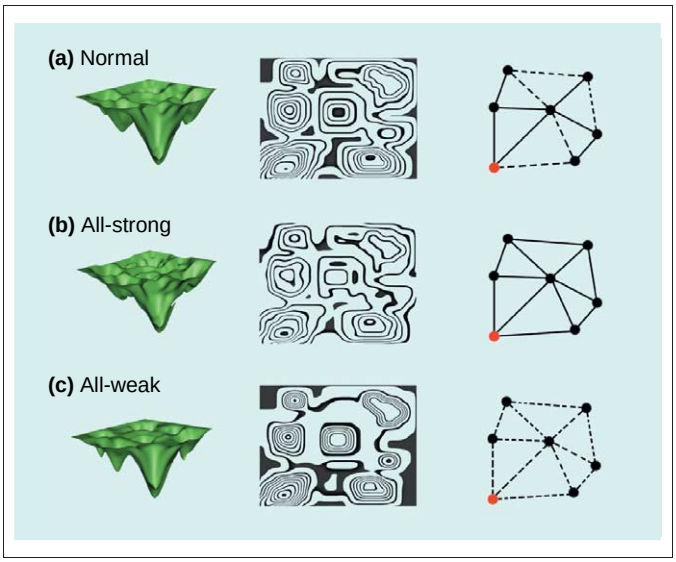

Figure 1. The importance of weak links for network stability. The 2D conformational space of a hypothetical protein is depicted in a **(a)** 'normal', **(b)** 'all-strong' and **(c)** 'all-weak' link situation. Energy is shown on the *z* axis. All seven energy minima remain the same in the energy landscapes in (a), (b) and (c). However, the activation energies between the minima are variable, minimal and maximal in (a), (b) and (c), respectively. For each, a 3D representation of the energy surface (green), contour plots of the energy levels around the seven minima (middle panels) and network representations of the transitions between the minima (drawn using the concept described in Ref. [20]) are shown. In the network representations, the red spheres represent the absolute energy minimum, black spheres represent local energy minima, unbroken lines represent strong links and broken lines represent weak links.

distribution of link strength is balanced between strong and weak links. The importance of weak links in 'smoothing' the energy landscape of protein folding is highlighted by recent findings that water has a decisive role in helping protein folding [21], confirming earlier assumptions [22].

Concluding remarks

The examples of social, ecosystem and cellular networks that are highlighted here show that an optimal balance between strong and weak links is required: strong links define the system, but, for system stability, many weak links might be needed to avoid unbearable fluctuations (a kind of noise-based error catastrophe) or decomposition of the network. It will be an exciting challenge for future studies to show the weak-link-dependent stabilization of organs, cells and probably even molecules or particles as networks.

Acknowledgements

I thank Robin Palotai and Máté Szalay (members of the High School Research Student Movement; www.kutdiak.hu) for help in the preparation of Figure 1. In addition, the advice of the two anonymous referees is greatly acknowledged. Work in the author's laboratory is supported by research grants from the EU (FP6506850), Hungarian Science Foundation (OTKA-T37357) and Hungarian Ministry of Social Welfare (ETT-32/03).

References

1 Watts, D.J. and Strogatz, S.H. (1998) Collective dynamics of 'small-world' networks. *Nature* 393, 440–442
2 Barabasi, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science* 286, 509–512
3 Newman, M.E.J. (2003) The structure and function of complex networks. *SIAM Rev.* 45, 167–256
4 Sole, R.V. *et al.* (2003) Selection, tinkering and emergence in complex networks. *Complexity* 8, 20–33
5 Almaas, A. *et al.* (2004) Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* 427, 839–843
6 Edwards, J.S. *et al.* (2001) *In silico* prediction of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* 19, 125–130
7 Mushegian, A.R. and Koonin, E.V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10268–10273
8 West, G.B. *et al.* (2002) Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2473–2478
9 Dodds, P.S. *et al.* (2001) Re-examination of the '3/4 law' of metabolism. *J. Theor. Biol.* 209, 9–27
10 Arita, M. (2004) The metabolic world of *Escherichia coli* is not small. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1543–1547
11 Goh, K-I. *et al.* (2001) Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.* 87, 278701
12 Bak, P. (1996) *How Nature Works. The Science of Self-Organized Criticality*, Springer-Verlag
13 Kauffman, S.A. (2000) *Investigations*, Oxford University Press
14 Rutherford, S.L. and Lindquist, S. (1998) Hsp90 as a capacitor for morphological evolution. *Nature* 396, 336–342
15 Vogelstein, B. *et al.* (2000) Surfing the p53 network. *Nature* 408, 307–310
16 Maslow, S. and Sneppen, K. (2002) Specificity and stability in topology of protein networks. *Science* 296, 910–913
17 Edelman, G.M. and Gally, J.A. (2001) Degeneracy and complexity in biological systems. *Proc. Natl. Acad. Sci. U. S. A.* 98, 13763–13768
18 Derenyi, I. *et al.* (2004) Topological phase transitions of random networks. *Physica A* 334, 583–590
19 Rao, C.V. *et al.* (2002) Control, exploitation and tolerance of intracellular noise. *Nature* 420, 231–237
20 Doye, J.P.K. (2002) The network topology of a potential energy landscape: a static scale-free network. *Phys. Rev. Lett.* 88, 238701
21 Levy, Y. and Onuchic, J.N. (2004) Water and proteins: A love-hate relationship. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3325–3326
22 Csermely, P. (1999) The 'chaperone-percolator' model: a possible molecular mechanism of Anfinsen-cage type chaperone action. *Bioessays* 21, 959–965
23 Silk, J.B. *et al.* (2003) Social bonds of female baboons enhance infant survival. *Science* 302, 1231–1234
24 Granovetter, M. (1973) The strength of weak ties. *Am. J. Sociol.* 78, 1360–1380
25 Degenne, A. and Forse, M. (1999) *Introducing Social Networks*, SAGE Publications, London, UK
26 Dodds, P.S. *et al.* (2003) Information exchange and the robustness of organizational networks. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12516–12521
27 McCann, K.S. *et al.* (1998) Weak trophic interactions and the balance of nature. *Nature* 395, 794–798
28 Berlow, E.L. (1999) Strong effects of weak interactions in ecological communities. *Nature* 398, 330–334
29 Edwards, J.S. and Palsson, B. (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U. S. A.* 97, 5528–5533

0968-0004/$ - see front matter
doi:10.1016/j.tibs.2004.05.004