Ann. Phys. (Berlin) **18**, No. 12, 822 – 829 (2009) / **DOI** 10.1002/andp.200910378

Network modules help the identification of key transport routes, signaling pathways in cellular and other networks

Robin Palotai* and **Peter Csermely****

Semmelweis University, Department of Medical Chemistry, Tuzolto str. 31-41, 1094 Budapest, Hungary

Received 1 September 2009, accepted 22 September 2009
Published online 11 December 2009

Key words Networks, learning, signal transduction, transport.
PACS 89.75.Fb, 89.75.Hc

Complex systems are successfully reduced to interacting elements via the network concept. Transport plays a key role in the survival of networks – for example the specialized signaling cascades of cellular networks filter noise and efficiently adapt the network structure to new stimuli. However, our general understanding of transport mechanisms and signaling pathways in complex systems is yet limited. Here we summarize the key network structures involved in transport, list the solutions available to overloaded systems for relaxing their load and outline a possible method for the computational determination of signaling pathways. We highlight that in addition to hubs, bridges and the network skeleton, the overlapping modular structure is also essential in network transport. Path-lenghts in the module-space of the yeast protein-protein interaction network indicated that module-based paths may cross fewer modular boundaries than shortest paths. Moreover, by locating network elements in the space of overlapping network modules and evaluating their distance in this 'module space', it may be possible to approximate signaling pathways computationally, which, in turn could serve the identification of signaling pathways of complex systems. Our model may be applicable in a wide range of fields including traffic control or drug design.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

1 Introduction

The network concept is successfully applied to reduce complex systems into a set of interacting elements connected by links to examine, understand and predict the topology, dynamics and emergent properties of the system [1–4]. In most networks the elements are autonomous agents, which not only develop direct interactions via links, but also establish long-range indirect interactions via various transport processes. The necessity of a transport process is usually evoked by a need for communication (social and telecommunication networks [5, 6]), transfer of resources (logistic networks, power grids [7, 8]) or regulation ensuring a fast, magnified and efficient response (signal transduction networks [9, 10]). The transport process is not only an emergent property of the network, but also a significant organizing force behind the structure and dynamics of the network. Links of the network may emerge in order to serve the transport process [11] and disappear later, if their existence is not required anymore or becomes even harmful [12].

Network signaling may be considered as a highly specialized case of transport. Signaling of cellular networks is a system level response to an incoming stimulus and is an extremely selective behavior fine-tuned by evolution. It efficiently filters noise-like stimuli, while quickly develops complex signaling cascades in response to a recognized stimulus [13, 14].

How does a network learn to discriminate between signal and noise? Our own studies [14–17] may help us to describe a common scenario: when an unusual signal arrives, which is strong and persistent

* E-mail: palotai.robin@gmail.com
** Corresponding author E-mail: csermely@eok.sote.hu, Phone: +36 1 459 1500 / 60130 Fax: +36 1 266 2650

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

enough to modify network behavior, the network slightly or profoundly disassembles: as a major process network modules (groups, communities [3, 18]) become loosely attached with a decreased overlap. When the stimulus is over, the network reassembles again. In this phase a large number of inter-modular contacts become re-established. However, these inter-modular contacts will not be exactly the same as before the stimulus: by developing a structural 'imprint' of the signal, the complex system has now a memory, it learned, on one hand which links may be more effective to dissipate the stimulus most efficiently, and on the other hand, which links are disturbing this process. If a similar stimulus arrives regularly (or the stimulus is large enough that all networks which were unable to learn the reorganization described above will disassemble and die) than the newly selected pathway may become dominant and may behave as a signaling pathway from then on.

In Sect. 2 we enumerate the main properties differentiating transport processes and consider optimality criteria with an emphasis on network throughput, then summarize the basic structures utilized by the network for efficient transport and filtering, namely the network skeleton, hubs (highly connected elements), bridges (elements connecting sparsly inter-connected network segments) and network modules (communities). In Sect. 3 we propose a method for reconstructing simulated and signaling pathways based on overlapping network module information. In Sect. 4 we summarize our findings and conclude.

2 Characterization of network transport

As described in Table 1 summarizing the main properties differentiating transport processes, the goal of the transport process is usually related to the survival of the system defining the network and therefore transport is related to the survival of the network as a connected graph with a large giant component. Transport may mobilize resources or information between network elements which are then used for the benefit of the system described by the network. For example, a network element in need may propagate a request message and other elements may send resources in response – this signaling scenario could emerge without network elements having attributable intentions or desires. 'Transport-provoking' cooperation may emerge through evolutionary mechanisms like signaling games [19]. The recently introduced protein games might also play a similar role in case of amino-acid networks [17, 20].

Table 1 Main properites of transport processes

Property	Description
Purpose	Goal of the transport process, usually related to the fitness or survival of network elements and the network as an entity.
Sources and sinks	Specific network elements may be identified as the source or sink of a given transported quantity.
Information need	Routing mechanism of network elements determining which neighboring element will they forward a received quantity to may require either local, global or intermediate (mesoscopic) knowledge about the network.
Determinism	Routing may be deterministic or stochastic.
Adaptiveness	Routing may (adaptive process) or may not (static process) be affected by the dynamic properties of the network.
Information preservation	Quantities transported may remain unchanged, suffer distortion or even get lost.
Time	Transport may be a discrete- or continous-time process.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

As for routing, local mechanisms tend to be stochastic because they lack extended information and therefore cannot be certain about the effectiveness of any single deterministic choice, while informed global routing mechanisms usually favor more deterministic approaches. Cellular networks exhibit local routing property, as the transmitted signal can be represented by the propagation of conformational changes of interacting proteins, and such changes may be evaluated locally via means of induced fit, conformational selection or protein games [17, 21, 22].

Let us take a detour and examine the optimality criterion of Table 2, which lists common expectations originally set towards computer networks [23]. It should be noted that an universally applicable set of optimality criteria does not exist, mainly due to that different optimality criteria are usually in conflict with each other. Conventional optimality criteria include low duration of delivery, short delivery paths or high transport throughput. In Box 1 we summarize a simple, yet descriptive model of network transport, which is sufficiently abstract not to distract attention with implementation details, but still lets us draw conclusions about the network transport processes focusing on the criteria mentioned above [11, 24–26].

Table 2 Expectations toward transport processes

Expectation	Description
Soundness	Operation of the transport process should strive to achieve and maintain its goal.
Simplicity	Among transport processes of similar performance the one with the simplest mechanism is preferred.
Robustness	The transport process should resist network failures, or at least degrade gracefully, proportional to network load or damage.
Stability	Operation of the transport process should lead to an equilibrium state of the network under stable circumstances.
Fairness	The transport process should strive to satisfy the transportation needs of all network elements equally.
Optimality	Operation of the transport process should be optimal for a set of criteria.

For example, in cell signaling networks short delivery paths would be preferred to reduce the distortion or loss of information (and indeed, most cellular networks are small-worlds [2, 4, 8]), while higher throughput would let the signaling network handle more stimuli simultaneously. Unfortunately shorter delivery paths increase the load on network elements of high centrality, and this, in turn, lowers the maximum possible throughput. As the example described here shows a transport process may conform to different optimization criteria to some extent but not all of them simultaneously. Moreover, if the satisfaction of multiple optimization criteria involves an increased complexity of the transport process, this increased complexity may hurt our expectations of simplicity, robustness or stability.

Knowing that the topology of any network sets an implicit upper bound on the maximum possible network throughput [27], it is interesting to investigate what kind of measures could the network utilize – apart from rearranging or coarsening its link structure [12, 28, 29] – in order to relax overloaded elements and prevent congestion.

First, network elements may exhibit adaptive behavior of taking into account the load of other elements in their routing mechanism. This behavior is exemplified by the multiple copies of protein isoforms in critical positions of cellular networks, such as the ‘critical nodes’ defined by Kahn and co-workers [30].

Second, network elements may resist to transport more than a given quantity, resulting in a filtered, faulty transport process but relaxing the load on the elements of the network. It is not surprising that network elements and structures of high centrality (having consequently a high load) are natural candidates

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Box 1 A simple model of network transport

In the original model of [11] information packets denoted a^{ik} are traveling from the source network element i to the sink network element k in the network $G = (N, E)$. In each discrete timestep first $R = \rho N$ new packets a^{ik} are generated with i and k chosen randomly and are added to the pool of packets at i denoted Q_i. Then for each element u a count of packets $C_u = C$ are randomly removed from Q_u and forwarded according to the routing strategy, or to the sink k if it is neighboring u (and thus the packet is removed from the network). Fig. 1 shows an example scenario.

This discrete transport model is very flexible: First, both sources and sinks can be identified, however this is not strictly necessary for its application. Second, depending on the applied routing strategy, the information need may be either local, global or mesoscopic, transport may either be stochastic or deterministic, information may be preserved or lost with some probability at each step of routing (see Table 1).

If the routing strategy is static and Markovian (the packet routing is independent of previously visited network elements), then both the expected path length and the expected load B_u of any element u (called *effective betweenness*) can be analytically derived. Moreover, if we define congestion as a state where exists an element u in the network with the packet pool Q_u growing faster than the processing capacity C_u of that element, then the throughput of the transport process can be characterized with the maximum $R = R_c$ value without congestion, given by $R_c = \min\{C_u N(N-1)/B_u\}$. Note that if any $C_u = C$ then the network throughput is capped by the element of maximum effective betweenness.

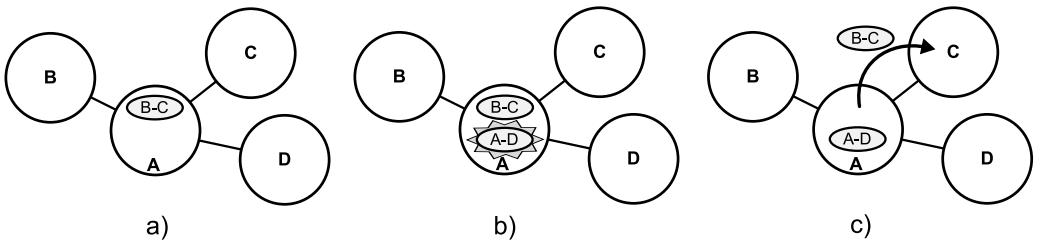

Fig. 1 An example scenario for the simple model of Box 1. a) The network element A and its neighboring elements B, C and D are shown. The packet pool of element A is shown containing a packet denoted B–C with source element B and sink element C. b) Generation step: The new packet A–D is generated at element A with random sink element of D. c) Transport step: A random packet (now packet B–C) is selected from the packet pool of A and becomes forwarded according to the routing mechanism. As packet B–C reaches its sink element C, it will disappear from the network.

for such filtering, because these elements of high centrality are expected to constitute a network skeleton or superhighway of transport [31] and thus are able to filter excessive amounts of transported quantities. Generally, congestion affects most the communication boundaries, such as central hubs of hierarchical networks or overlaps of network modules, both providing bridges between different network segments [32]. If we define modules as having more intra-modular links than inter-modular [33], then modules themselves also act as noise traps with noise rather circulating inside the module and eventually getting dissipated instead leaving the module.

Third, the routing mechanism may decide to sacrifice certain optimization criteria in favor of network throughput by deliberately utilizing alternative or back-up routes to some extent in parallel with the network skeleton. This procedure is not necessary adaptive, for example (overlapping) network module information may serve as a basis for static routing if known [34, 35].

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

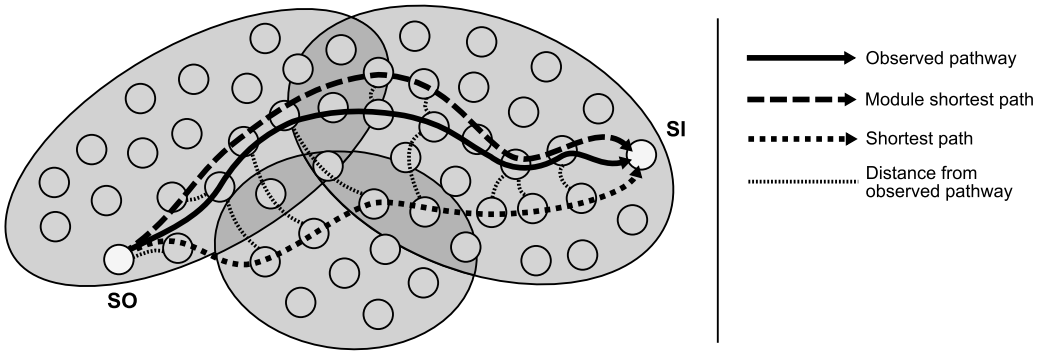

Fig. 2 An illustrative network with three overlapping modules (marked with ellipses) is shown. Small circles denote network elements, links are not visible. One experimentally observed and two simulated pathways connecting a source element (SO) with a sink element (SI) are shown: the actually observed pathway (continuous line), a module-based shortest path with distance metric calculated in module space as described in Sect. 3 (dashed line) and a traditional shortest path (dotted line). Thin dotted lines indicate the distance between elements of simulated pathways and the observed pathway. In the scenario of the illustrative figure, the module-based shortest path approximates the observed pathway better than the traditional shortest path.

3 Simulated pathways on overlapping modules

Recent advances in network module identification methods are not only able to assign elements to multiple overlapping modules but also provide metrics describing membership strength of any elements to different modules [18, 36–39], effectively locating network elements in the M-dimensional module space, M being the number of modules in the network. Therefore it is possible to evaluate a structural compatibility between network elements based on the distance of elements in the module space. This fact, combined with the observations mentioned in Sect. 2 that 1) propagation of transmitted signals in cellular networks can be evaluated via local compatibility metrics between network elements and 2) information on overlapping network modules may serve as a basis for routing, raises the question if, module-based simulated pathways between network elements are correlated to actual pathways experimentally observed in the network.

In order to decide, if the observed and simulated pathways are correlated, one may compare observed pathways p, traditional shortest paths $\hat{p}_{SP}$ and module-based shortest paths $\hat{p}_{MSP}$: in the latter case the distance $d^m_{u,v}$ of network elements u and v is the Manhattan distance $\sum_k |b_u[k] - b_v[k]|$, where b_i is a vector of M components with $b_i[k]$ being the membership strength of network element i to the module k. Membership strength is a modularization method-dependent measure characterizing the affinity of network element i to be part of the module k – for example, the sum of the components of the 'fuzzy membership degree' vector of a given element is unity in [38], while in the ModuLand framework a custom centrality measure (the landscape centrality) is calculated for each element, and this centrality value is distributed between different network modules in the ratio of the membership affinity of the given element to the respective modules [15, 37].

To compare an observed pathway p with a simulated pathway $\hat{p}$, one may calculate the distance $d(p, \hat{p})$ by summing the $d(u_{\hat{p}}, p)$ and $d(u_p, \hat{p})$ distances between elements $u_q \in q$ and the respective pathway, where $d(u_{q_1}, q_2)$ is the minimum distance between element u_{q_1} and any element $w \in q_2$, calculated by Dijkstra's algorithm of shortest paths. Finally the normalized $d'(p, \hat{p})$ is introduced as $d(p, \hat{p})/\,|p|$, where $|p|$ denotes the number of elements in p. Fig. 2 shows an illustrative scenario of pathway comparison.

If $d'(p, \hat{p}_{MSP})$ would prove to be generally lower than $d'(p, \hat{p}_{SP})$, then we could conclude that module-based shortest paths are better approximators of module-dependent network pathways than shortest paths.

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

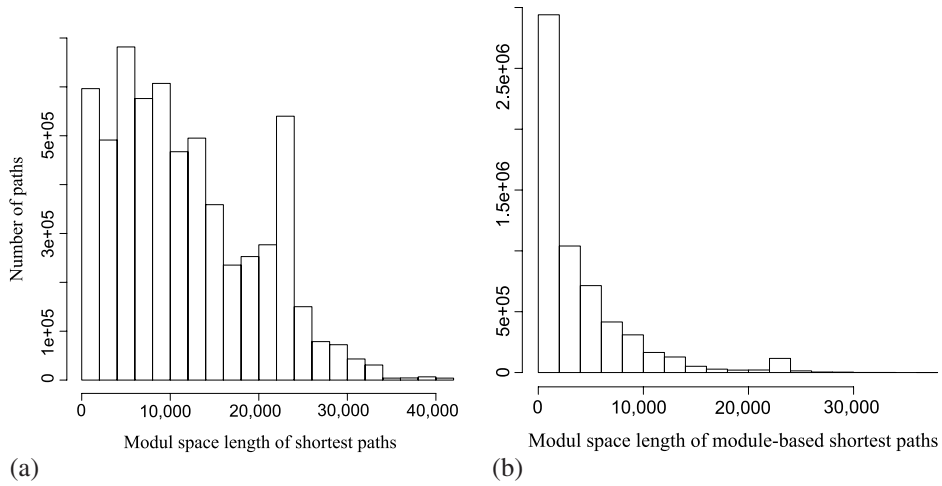

Fig. 3 Histogram of the module-space lengths of **a)** shortest paths and **b)** module-based shortest paths between all pairs of the 2,444 network elements of the main component of the protein-protein interaction network of *S. cerevisiae* [40]. The significant difference of distributions ($p < 0.01$, Wilcoxon matched pairs signed rank test) means that traditional shortest paths are not shortest paths in the module-space, possibly crossing more network module boundaries than the possible minimum. Overlapping modular structure of the network was uncovered using the NodeLand and TotalHill procedures of the ModuLand modularization method [37] and module-space length of paths were calculated by evaluating the distance of network element pairs via the $d^m_{u,v}$ metric (see Sect. 3), where the ModuLand landscape centrality of a given element is distributed between the components of the module membership strength vector of the given element in the ratio of the membership affinity of the given element to the respective modules. For each pair of elements a single shortest path was randomly chosen.

As a preliminary investigation we applied a reverse approach, and compared the module-space lengths of the shortest paths and the module-based shortest paths between all pairs of network elements of the protein-protein interaction network of the yeast, *S. cerevisiae*, as compiled in [40], via evaluating the distance between elements u and v using the $d^m_{u,v}$ measure introduced above, and calculated based on the overlapping modular structure of the network as uncovered by the NodeLand and the TotalHill procedures of the ModuLand modularization method described in [37].

Naturally, the module-space length of a module-based shortest path between elements i and j is a lower bound for the module-space length of any shortest path between the said elements, but Fig. 3 shows that the distribution of module-space lengths of module-based shortest paths gets concentrated on significantly lower values than the distribution of module-space lengths of shortest paths. This means that traditional shortest paths are not shortest paths in the module-space, possibly crossing more network module boundaries than the possible minimum, which in part explains the negative impact of traditional shortest paths on network throughput as described in Sect. 2. It must be noted that our comparison did not take all possible shortest paths between a given pair of network elements into account. Instead, for each pair of elements a single shortest path was randomly chosen, which could introduce a sampling bias into the presented result. This uncertainty remains yet to be resolved in subsequent studies.

Modules often correspond to various functions of the system coded by the network. Therefore, the modular analysis described above may also help us to determine key signaling pathways as inter-modular routes. This becomes especially likely, if we take into account the hierarchical structure of modules, where modules of the original layer are represented as elements of the next layer of hierarchy [35, 37].

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

4 Summary

We have described that transport processes are significant organizing forces of the network structure and dynamics, and considered a mechanism of network signaling filtering noise and adapting to newly recognized stimuli. We investigated the main properties differentiating transport processes, listed expectations towards transport processes and noted that an universally applicable set of optimization criteria does not exist due to criterion-conflicts. We examined the limits of optimizing network transport for highest possible throughput in the framework of a simple, yet descriptive model of network transport and described the ways how different network structures could cause, and, interestingly, also relax congestion.

We highlighted the role of overlapping network modules and proposed that exploiting the information on overlapping modules, for example the distance between network elements in the 'module space', may help the analysis of routing mechanisms. Finally, we asked the question if module-based simulated pathways between network elements are correlated to real pathways observed in the network, and suggested a method for determining the answer. Our preliminary results in the yeast protein-protein interaction network suggested that module-based pathways may cross fewer modular boundaries than shortest paths, which may have a significant impact in stressed and/or noisy cells and other complex systems.

If module-based pathways would describe well the real, observed pathways, the identification of key, signaling pathways of complex systems would become possible using higher layers of the hierarchical modules. Such knowledge could be utilized in a wide range of fields including traffic control or drug design.

Acknowledgements Work in the authors' laboratory was supported by the EU (FP6-518230) and the Hungarian National Science Foundation (OTKA K69105).

References

[1] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079–1187 (2002).
[2] A. L. Barabási and Z. N. Oltvai, Nat. Rev. Genet. **5**, 101–113 (2004).
[3] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. U. Hwang, Phys. Rep. **424**, 175–308 (2006).
[4] P. Csermely, Weak links: Stabilizers of complex systems from proteins to social networks (Springer, Berlin, 2006).
[5] R. Radner, Econometrica **61**, 1109–1146 (1993).
[6] J. P. Scott, Social Network Analysis: A Handbook (Sage Publications, London, 2000).
[7] H. Bast, S. Funke, P. Sanders, and D. Schultes, Science **316**, 566 (2007).
[8] D. J. Watts and S. H. Strogatz, Nature **393**, 440–442 (1998).
[9] X. Zhu, M. Gerstein, and M. Snyder, Genes Dev. **21**, 1010–1024 (2007).
[10] E. Almaas, J. Exp. Biol. **210**, 1548–1558 (2007).
[11] R. Guimerà, A. Àrenas, A. Díaz-Guilera, F. Vega-Redondo, and A. Cabrales, Phys. Rev. Lett. **89**, 248701 (2002).
[12] A. E. Motter, Phys. Rev. Lett. **93**, 098701 (2004).
[13] H. Yu and M. Gerstein, Proc. Natl. Acad. Sci. **103**, 14724–14731 (2006).
[14] M. S. Szalay, I. A. Kovács, T. Korcsmáros, C. Böde, and P. Csermely, FEBS Lett. **581**, 3675–3680 (2007).
[15] R. Palotai, M. S. Szalay, and P. Csermely, IUBMB Life **60**, 10–18 (2008).
[16] A. Mihalik, R. Palotai, and P. Csermely, Biochemistry (Hung.) **32**, S67 (2008).
[17] M. A. Antal, C. Böde, and P. Csermely, Curr. Protein. Pept. Sci. **10**, 161–172 (2009).
[18] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature **435**, 814–818 (2005).
[19] A. Zahavi, J. Theor. Biol. **53**, 205–214 (1975).
[20] I. A. Kovács, M. S. Szalay, and P. Csermely, FEBS Lett. **579**, 2254–2260 (2005).
[21] D. E. Koshland, Proc. Natl. Acad. Sci. USA **44**, 98–104 (1958).
[22] G. Careri, P. Fasella, and E. Gratton, CRC Crit. Rev. Biochem. **3**, 141–64 (1975).
[23] A. S. Tanenbaum, Computer networks: 2nd edition (Prentice-Hall, Upper Saddle River, NJ, USA, 1988).

© 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

[24] B. H. Wang and T. Zhou, J. Kor. Phys. Soc. **50**, 134–141 (2007).
[25] G. Yan, T. Zhou, B. Hu, Z. Q. Fu, and B. H. Wang, Phys. Rev. E **73**, 046108 (2006).
[26] B. Tadic, G. J. Rodgers, and S. Thurner, Int. J. Bifurcat. Chaos **17**, 2363–2385 (2007).
[27] S. Sreenivasan, R. Cohen, E. López, Z. Toroczkai, and H. E. Stanley, Phys. Rev. E **75**, 036105 (2007).
[28] A. E. Motter, N. Gulbahce, E. Almaas, and A. L. Barabási, Mol. Syst. Biol. **4**, 168 (2008).
[29] T. Nishikawa, N. Gulbahce, and A. E. Motter, PLoS Comput. Biol. **4**, e1000236 (2008).
[30] C. M. Taniguchi, B. Emanuelli, and R. C. Kahn, Nat. Rev. Mol. Cell. Biol. **7**, 85–96 (2006).
[31] Z. Wu, L. A. Braunstein, S. Havlin, and H. E. Stanley, Phys. Rev. Lett. **96**, 148702 (2006).
[32] D. Gfeller, J. C. Chappelier, and P. De Los Rios, Phys. Rev. E **72**, 056135 (2005).
[33] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Proc. Natl. Acad. Sci. USA **101**, 2658–2663 (2004).
[34] L. Danon, A. Àrenas, and A. Díaz-Guilera, Phys. Rev. E **77**, 036103 (2008).
[35] R. Palotai, MSc. Thesis, Budapest University of Technology and Economics, Budapest, Hungary (2008).
[36] J. Reichardt and S. Bornholdt, Phys. Rev. Lett. **93**, 218701 (2004).
[37] I. A. Kovács, M. S. Szalay, P. Csermely, and T. Korcsmáros, Method for analyzing the fine structure of networks, Patent application number: WO2007093960 (2006).
[38] T. Nepusz, A. Petróczi, L. Négyessy, and F. Bazsó, Phys. Rev. E **77**, 016107 (2007).
[39] S. Zhang, R. S. Wang, and X. S. Zhang, Phys. A **374**, 483–490 (2007).
[40] D. Ekman, S. Light, A. Björklund, and A. Elofsson, Genome. Biol. **7**, R45 (2006).

 © 2009 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim