# Degree landscapes in scale-free networks

Jacob Bock Axelsen,[1,*] Sebastian Bernhardsson,[2] Martin Rosvall,[2] Kim Sneppen,[3] and Ala Trusina[3]
[1]*NBI, Blegdamsvej 17, Dk 2100, Copenhagen, Denmark*
[2]*Department of Theoretical Physics, Umeå University, 901 87 Umeå, Sweden*
[3]*NBI, Blegdamsvej 17, DK-2100, Copenhagen, Denmark*

(Received 16 December 2005; revised manuscript received 14 August 2006; published 26 September 2006)

We generalize the degree-organizational view of real-world networks with broad degree distributions in a landscape analog with mountains (high-degree nodes) and valleys (low-degree nodes). For example, correlated degrees between adjacent nodes correspond to smooth landscapes (social networks), hierarchical networks to one-mountain landscapes (the Internet), and degree-disassortative networks without hierarchical features to rough landscapes with several mountains. To quantify the topology, we here measure the widths of the mountains and the separation between different mountains. We also generate ridge landscapes to model networks organized under constraints imposed by the space the networks are embedded in, associated to spatial or in molecular networks to functional localization.

DOI: 10.1103/PhysRevE.74.036119 PACS number(s): 89.75.Hc, 87.16.Xa, 87.16.Yc, 89.75.Fb

## I. INTRODUCTION

The broad degree distribution in many real-world networks [1–4] makes it meaningful to investigate the topological organization of nodes in terms of their degree. It has been found that many social networks are assortative, with correlated degrees of adjacent nodes, but that technological and biological networks often are disassortative, with anticorrelated degrees of adjacent nodes [5–8]. The degree correlation profile, generated by comparison between the network and its randomized counterparts without degree correlations, uncovers in the Internet an over-representation of links between intermediate- and low-degree nodes and a slight over-representation of links between the nodes of highest degrees [9]. Contrary, most biological networks have a suppression of links between the hubs [10]. To characterize the organization beyond correlations between adjacent nodes, Trusina *et al.* [11] introduced the hierarchy measure $\mathcal{F}$. $\mathcal{F} \in (0,1]$, is the fraction of shortest paths between all pairs of nodes that are degree hierarchical [12]. A degree hierarchical shortest path has node degrees sorted monotonously or in an ascending order followed by a descending order. It was found that biological networks with decentralized hubs stand out from other networks with a very low value of $\mathcal{F}$.

Here we generalize the presented findings in a landscape analog, with mountains (high degree nodes) and valleys (low degree nodes). Qualitatively, the landscape analog can be considered as a mapping of constraints to move in a network, since the altitude can be interpreted as the density of random walkers on these nodes (with nodes as states and links as transition possibilities between different states [13]). With this interpretation, social networks form smooth landscapes without separated mountains, like the Internet with a single mountain with first ascending and then descending hierarchical paths, whereas biological networks form rough landscapes with several separated mountains and broken hierarchical paths.

To quantify the constraints imposed by the topology beyond correlations between adjacent nodes, we in this paper also measure the typical width of individual mountains and the separation between different mountains (Fig. 1). In this way we bring the landscape analog to a meaningful quantification of network topologies.

We also suggest a method to generate *ridge landscapes* [Fig. 2(c)]. In its simplest implementation, we assign a random rank to every node in a network, and organize the nodes hierarchically based on their rank. This method creates nonrandom networks, distinguished by a separation of hubs (disassortative with low $\mathcal{F}$). We argue that error proned ridge landscapes can represent networks organized under different spatial constraints put on real-world networks during their evolution.

## II. BACKGROUND ON DEGREE CORRELATIONS AND WALKS ON NETWORKS

The fact that the degree distribution in many real-world networks follows a broad distribution [1], with a significant number of nodes having very many links, opens for investigation of the topological organization of nodes in terms of their degree. This investigation becomes meaningful when

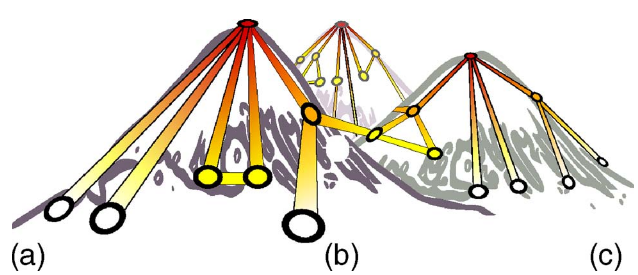


FIG. 1. (Color online) Degree landscape of a network with mountains and valleys, with the altitude of a node proportional to its degree. A route over one mountain corresponds to making a degree-hierarchical path [(a) to (b)] while climbing over more than one mountain breaks the degree-hierarchical path [(a) to (c)].

*Electronic address: bock@nbi.dk

 

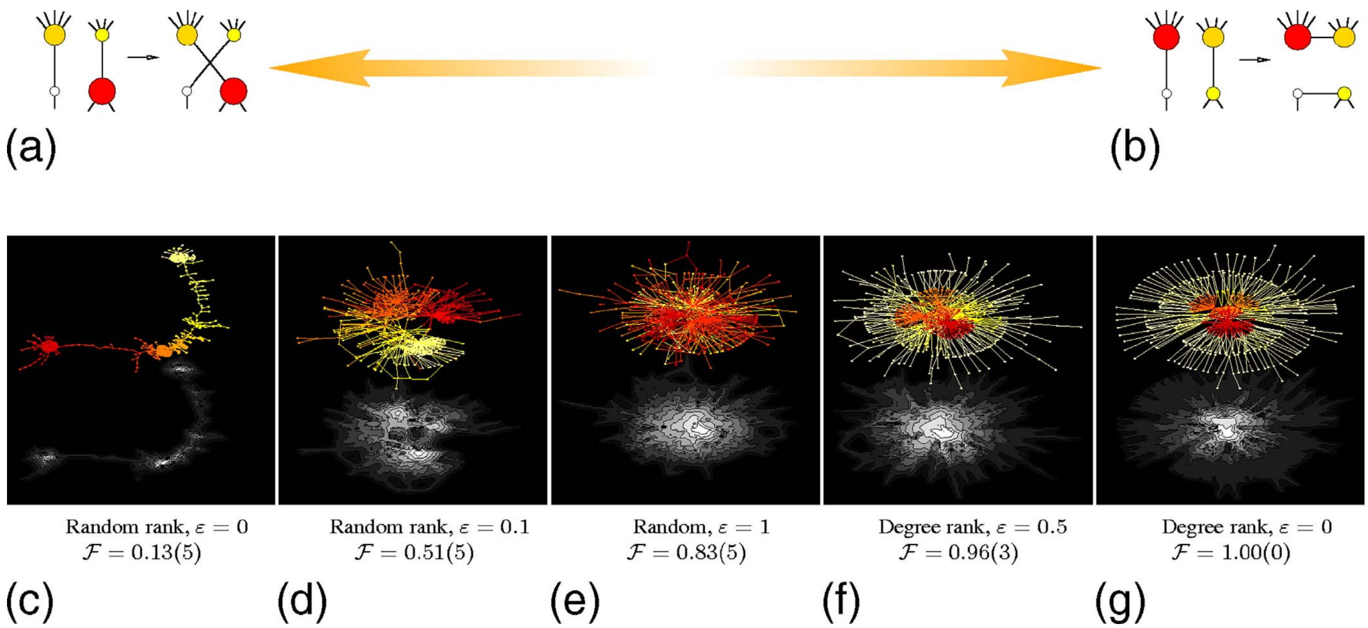


FIG. 2. (Color online) Visualization of degree landscapes of networks organized from ridge landscapes (c), via random landscapes (e), to peaked one-mountain landscapes (g). The links are pairwise swapped to connect high-ranked nodes to organize the nodes globally according to their rank (color coded from dark red for high rank, to light white for low rank), with random swaps at different rates $\varepsilon$. The rank is set randomly to the nodes, as in the swap example in (a), in (c) and (d), and proportional to the degree of the nodes, as in the swap example in (b), in (f) and (g). The random network in (e) corresponds to $\varepsilon = 1$. The corresponding visualization of the degree landscapes are color coded according to altitude from black (low) to white (high). The networks are scale-free with an exponent $\gamma = 2.5$ and of size $N = 400$, originally generated with the algorithm suggested [11]. The layout is generated with the Kamada-Kawai algorithm in Pajek [14].

the degree of a node is a proxy for the function of the node in the network [15].

For example, a natural question to ask is whether nodes of similar degree are or are not connected with each other. To answer this question Newman introduced the assortativity measure [5–7]. A network is said to show assortative mixing if nodes with many links tend to be connected to other nodes with many links in the network. Oppositely, if the nodes with many links tend to be connected to nodes with few links, the network is said to show disassortative mixing. It was found that several types of social networks of collaborations are assortative, and that technological and biological networks tend to be disassortative.

To be able to answer how the degrees in a network are correlated in more detail, Maslov and Sneppen introduced the degree-correlation profile [10]. An ensemble of *null models* are generated by randomizing the neighborhoods while keeping the point properties of the nodes. More precisely, if the degree of node $i$ is $k_i$ then this number is kept constant, but the neighbors of $i$ are exchanged. The algorithm obtains this by repeatedly taking two randomly chosen links and swap one node on the first link with a node on the second link. The correlation matrix is then calculated by taking the ratio of the probability of $k_i$-degree nodes having friends of degree $k_j$ in the observed network to the probability given by the average of the randomized ensemble: $P(k_i,k_j)/P_R(k_i,k_j)$. In this way it was identified that biological networks display a negative correlation of nodes of high degree with themselves, and a positive correlation between high and low degree nodes. Contrary, the Internet showed an over-representation of links between intermediate- and low-degree nodes and a slight over-representation of links between the nodes of highest degrees.

So far we have discussed local correlations. However, the position of a node in a network is not solely defined by its immediate neighbors. For example, communication in networks can extend to the whole network, like in the Internet. Therefore it also makes sense to ask which role the organization of the nodes in terms of their degrees plays in a global perspective. One way to attack this question is to study random walks as proxies for communication in the network.

Depending on the topology of the network, random walkers will either quickly visit nodes in distant parts of the network or perhaps stay for longer periods in local neighborhoods. However, the frequency at which a random walker will visit a node is proportional to the degree of the node [16]. The degree will therefore be a proxy for how central a node is in terms of communication in a network.

Eriksen *et al.* considered an approximation of this flow of random walkers, and showed that the Internet is highly modular with the U.S. military and the Russian military at the extreme ends [13].

Another way to approximate global signals in a network is to consider its hierarchical organization [17], and assume that signals follow hierarchical paths [11,12]. Consider only a shortest path $s$ between two nodes $i = s(1)$ and $l = s(n)$. Let the rank of node $i$ be $r_i$, proportional to the degree of node $i$, $r_i = k_i$, and study the degree sequence on the shortest path $\{k_{s(1)}, k_{s(2)}, \ldots, k_{s(n-1)}, k_{s(n)}\}$ from $s(1) = i$ to $s(n) = l$. The shortest path is said to be *degree hierarchical* if the node degrees are sorted monotonously or in an ascending followed by a descending order $(k_{s(1)} \leqslant \cdots \leqslant k_{s(a+1)} \geqslant \ldots \geqslant k_{s(a+b)}$

where $a+b=n$ and $a,b>0$). Moreover, the hierarchy measure $\mathcal{F}$ is the number of degree hierarchical shortest paths divided by the total number of shortest paths in the network. It was found that the Internet and a network of CEOs in the U.S. are highly hierarchical. Contrary, a protein-protein interaction network in yeast turned out to be highly antihierarchical when compared with its random counterpart [11]. This observation stimulated our search for a process responsible for reorganizing a network with a broad degree distribution into such a topology.

### III. RANK-HIERARCHICAL LINK SWAPPING

We here start by reviewing the method presented by Trusina *et al.* [11] to generate degree-hierarchical networks. The networks evolve by pairwise rewirings of the links, with every rewiring constrained by the rank of the nodes involved in the rewiring. The rank of a node is proportional to its degree in the *degree-rank hierarchy* [see Fig. 2(b)], and set to a random rank (degree independent) in the *random-rank hierarchy* [see Fig. 2(a)]. At every time step, two random links are chosen and reconnected such that the two nodes with the highest ranks become adjacent. In this way the degree of every node is kept constant and the nodes are globally organized in decreasing rank order.

To be able to investigate networks in between, respectively, the random-rank hierarchies, the degree-rank hierarchies, and random networks, we also allow for random link swaps without the constraints set by the rank of the nodes. To make a random link swap with probability $\varepsilon$ corresponds in this way to an error rate in the creation of the extreme networks. When $\varepsilon \to 1$ the methods become equivalent to the randomization of networks with remaining degree sequence suggested in Ref. [10], see Fig. 2(e).

Figure 2 shows topologies generated with the different models. They all originate from a random scale-free network [shown in Fig. 2(e)] with degree distribution $P(k) \propto k^{-2.5}$ and system size $N=400$, generated with the method suggested in Ref. [11]. The extreme networks, the perfect random-rank hierarchy in Fig. 2(c) and the perfect degree-rank hierarchy Fig. 2(g) ($\varepsilon=0$), surround the networks with increasing error rates toward the random scale-free network with $\varepsilon=1$ in the middle [Fig. 2(e)]. The perfect degree-rank hierarchy consists of a tightly connected core of large degree nodes that forms a very peaked mountain in the degree sequence of all shortest paths ($\mathcal{F}=1$). This property will lead most layout algorithms to place the hubs at the center of the figure. Here, we have employed the Kamada-Kawai algorithm of the program Pajek [14], which ensures that no two nodes are projected on top of each other.

The random-rank hierarchy, on the other hand, forms a very stringy and nonrandom structure—a ridge landscape. The length of the string is of the order $D \propto N$, with very long pathways that break the small-world property found in most real-world networks. However, as for the original small-world scenario proposed by Ref. [18], the large diameter of the stringy scale-free networks collapses if small perturbations exist in the hierarchical organization. If we generate the network with a small error rate $\varepsilon$, the diameter of the network collapses as seen in Fig. 2(d). Note that the color gradient indicates that the random-rank hierarchy is still intact at this stage, and that the hubs (local mountains) are separated. The degree-rank hierarchy in Fig. 2(f) is rewired with a higher error rate $\varepsilon=0.5$, while still maintaining a high level of hierarchical organization.

In both cases, the two organizing principles leads to higher clustering [18], with more loops of length 3 and longer [19], than in the random counterparts (not shown). The organization along an arbitrary coordinate tends to make friends of friends more alike up to the limit set by the width of the mountains.

In Fig. 2, we also quantify the degree-hierarchical organizations of the scale-free networks organized by, respectively, degree rank and number rank. For the random scale-free network with degree distribution $P(k) \propto k^{-2.5}$ and $N=1000$ nodes, $\mathcal{F}=0.83 \pm 0.05$. The networks organized hierarchically according to degree rank [as in Fig. 2(b)] have $\mathcal{F}=1$ as expected. Further, when introducing a finite error rate $\varepsilon$ for link rewirings toward the degree hierarchy we find that its topology is robust in the sense that both diameter (not shown) and $\mathcal{F}$ remain unchanged for even quite large errors. The perfect random-rank hierarchy has a much lower degree-hierarchical organization, $\mathcal{F}=0.13 \pm 0.05$. Because of the collapsing diameter, the random-rank hierarchy is not as robust as the degree-rank hierarchy to errors in the rewiring.

Figure 3 visualizes, in increasing degree-hierarchical order, the degree landscapes of a number of real-world networks: Yeast in (a) is the protein-interaction network in *Saccharomyces Cerevisia* detected by the two-hybrid experiment [20], Manhattan in (b) is the dual map of Manhattan with streets as nodes and intersections as links [21], and the Internet in (c) is the network of autonomous systems [22]. Internet and Manhattan consist of one single mountain with first ascending and then descending hierarchical paths, whereas yeast forms a rough landscape with several mountains and broken hierarchical paths.

### IV. MEASURING DEGREE LANDSCAPES

To bring the landscape analog to an insightful quantification of network topologies (independent of the means of visualization), we introduce two measures. First, inspired by the information horizons in networks [23–25], we present a revised hierarchy measure $\mathcal{F}(\ell)$, to estimate the size of the mountains. $\mathcal{F}(\ell)$ is the fraction of pairs of nodes at distance $\ell$ that are hierarchically connected. Figure 4(a) shows that $\mathcal{F}(\ell)$ decreases fastest for the random-rank hierarchy at a length scale $\ell \approx 4$ corresponding to the width of the ridge landscape shown in Fig. 2(c). Figures 4(b)–4(d) shows the real-world networks as in Fig. 3 compared with their random counterparts with the same degree sequence. Yeast behaves qualitatively like the random-rank hierarchy with $\varepsilon$ between 0 and 0.1, which probably reflects some functional localization. Contrary, despite their embedding in real space, the Internet and Manhattan both have a substantial fraction of long degree-hierarchical paths, corresponding to wide mountains. However, the randomized counterparts of the two latter networks, with more peaked mountain landscapes, are both

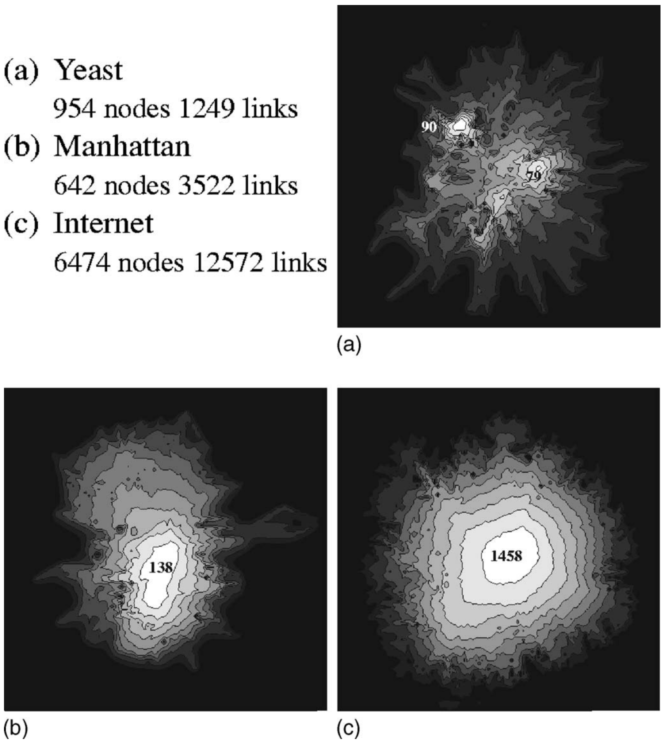


FIG. 3. Visualization of degree landscapes of real-world networks. The coloring of the altitudes are relative to the summit altitude. Yeast in (a) is the protein-protein interaction network in *Saccharomyces Cerevisia* [20], Manhattan in (b) is the dual map of Manhattan with streets as nodes and intersections as links [21], and the Internet in (c) is the network of autonomous systems [22]. The topological maps are not based on the real space the networks are embedded in, but the Kamada-Kawai algorithm in Pajek [14].

more degree hierarchical than the real networks.

We define the width of a mountain as the length where 50% of the paths are hierarchical. Figure 4 shows that the average width of the mountains in the random-rank hierarchy and yeast is about 4. In Manhattan and the Internet it is larger, about 6, and in the degree-rank hierarchy it is by definition the network diameter.

In the second landscape measure, we measure the separation between mountains to investigate how the hubs are positioned relative to each other. $d(k_{\text{hub}})$ is associated to maximum distances between nodes with degree $k$ equal to or larger than the threshold value $k_{\text{hub}}$. It is defined by the distance from one hub to its most distant hub in the network, averaged over all hubs

$$d(k_{\text{hub}}) = \frac{1}{N_{k \geq k_{\text{hub}}}} \sum_{\{i|k_i \geq k_{\text{hub}}\}} \max_{\{j|k_j \geq k_{\text{hub}}\}} d_{ij}, \qquad (1)$$

with $d_{ij}$ being the length of the shortest path between $i$ and $j$, and $k_i$ the degree of node $i$. The value of $d(k_{\text{hub}})$ is highly dependent of the definition of a hub, and we therefore measure $d(k_{\text{hub}})$ for all values of $k_{\text{hub}}$.

Figure 5 shows $d(k_{\text{hub}})$ for a few different networks and their random counterparts. Figure 5(a) shows that the one-mountain landscapes, the degree-rank hierarchy and the random network, both have hubs tightly connected. Contrary, the hubs in the perfect random-rank hierarchy are extremely separated $[d(1) \approx 100]$ all the way out to a very high hub-threshold value (curve not shown). All the real-world net-

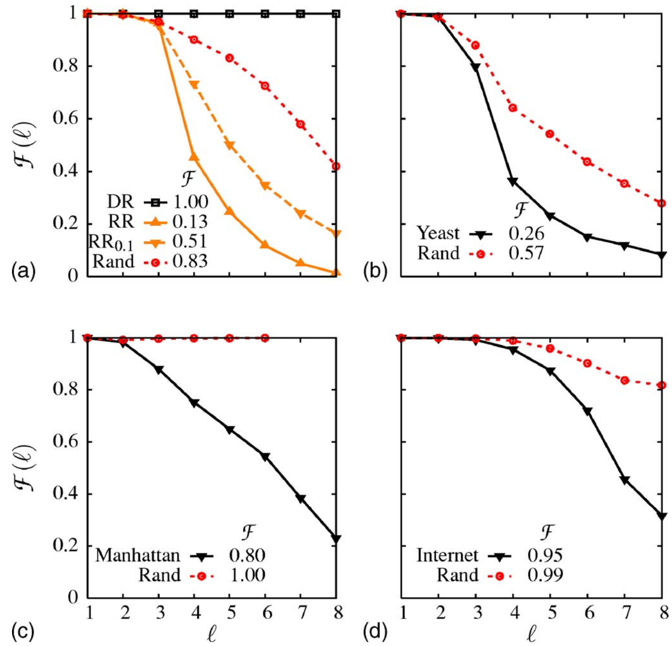


FIG. 4. (Color online) The degree-hierarchical organization as a function of path length. $\mathcal{F}(\ell)$ is the fraction of pair of nodes, separated by a distance $\ell$, that are connected by a degree-hierarchical path. (a) shows the two model networks: The degree-rank hierarchy (degree-rank), the random-rank hierarchy (random-rank) for $\varepsilon=0$ and 0.1, together with the random scale-free network (random). The real-world networks in (b)–(d) are the same as in Fig. 3. All networks are compared with their random counterparts (Rand).

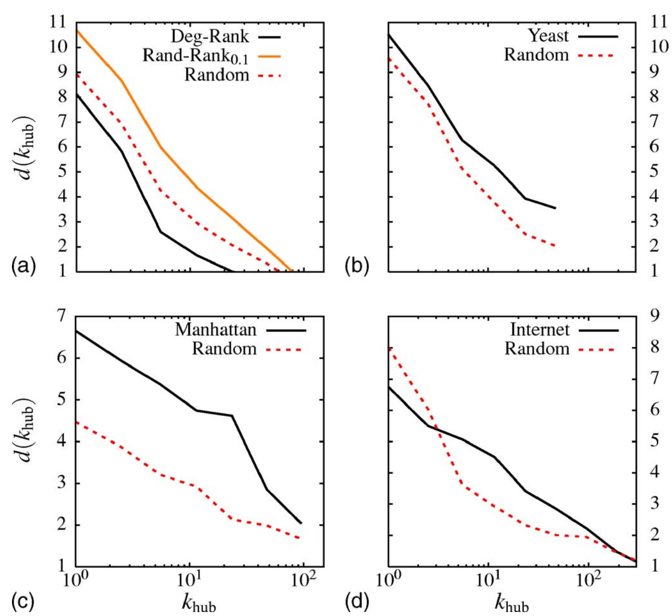


FIG. 5. (Color online) Hub separation in networks measured as the average longest distance $d(k_{\text{hub}})$ between all nodes of degree $k \geq k_{\text{hub}}$, Eq. (1), for the same networks as in Fig. 4. The results are $\log_2$ binned.

works in Figs. 5(b)–5(d) fall in between these extremes, but with a higher $d(k_{\mathrm{hub}})$ than randomly expected for most values of $k_{\mathrm{hub}}$. Manhattan [Fig. 5(c)] and the Internet [Fig. 5(d)] are close to random for high degrees, while yeast [Fig. 5(b)] has a separation for all sizes. The close resemblance between the random-rank hierarchy and yeast in Figs. 4 and 5 suggests that the separation of hubs probably reflects a separation of functions at all scales.

Manhattan is mainly a planned city where the largest hubs, corresponding to streets and avenues, are connected to each other in a bipartite way. This results in a $d(k_{\mathrm{hub}})$ close to 2 for the largest hubs. The Internet is constructed with a hierarchical structure within each country, and all intermediate-degree nodes (typically connected to low degree nodes [9]) are therefore separated from each other globally. However, the largest hubs interconnect the countries, and are therefore connected with each other. This results in a $d(k_{\mathrm{hub}})$ close to 1 for the largest hubs.

## V. CONCLUSION

To summarize, with the starting point that the hubs play an important role in the network function, we have asked how the hubs are positioned relative to each other in different networks. The approach was to generalize the degree-organizational view of real-world networks with broad degree distributions, illustrated in a landscape analog with mountains (high degree nodes) and valleys (low degree nodes). We have also generated ridge landscapes to model networks organized under constraints imposed by the space the networks are embedded in, associated to spatial or in molecular networks to functional localization. By ordering nodes associated to random numbers, we present a simple way of taking classification of nodes into account and the constraints this sets on the network topology.

To quantify this connection between function and topology in the illustrative landscape analog, and to be able to compare real-world networks, we have measured the widths of the mountains and the separation between different mountains. We found that the dual map of Manhattan consists approximately only of one mountain. This implies that typical navigation between a source and a target street in the city involves first going to larger and larger streets, and then to smaller and smaller streets. The Internet shares this one-mountain landscape, and a routed package will experience the same landscape. But, since the spatial constraints are weaker in the Internet, the width of the mountain is about the same as in the city despite the substantially larger network. Finally, the topological landscape in the protein-interaction network in yeast has a different topology with numerous separated hills, which probably corresponds to functional localization. This suggests that signals within the compartments dominate over global signaling.

## ACKNOWLEDGMENTS

We acknowledge the support from the Danish National Research Foundation through "Models of Life" at NBI, and the Lundbeck Foundation for funding Ph.D. studies.


[1] A. L. Barabasi and R. Albert, Science **286**, 509 (1999).
[2] R. Albert and A. L. Barabasi, Rev. Mod. Phys. **74**, 1 (2002).
[3] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[4] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[5] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[6] J. Park and M. E. J. Newman, Phys. Rev. E **68**, 026112 (2003).
[7] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).
[8] R. Pastor-Satorras, A. Vazquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[9] S. Maslov, K. Sneppen, and A. Zaliznyak, Physica A **333**, 529 (2004).
[10] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
[11] A. Trusina, S. Maslov, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. **92**, 178702 (2004).
[12] L. Gao, IEEE/ACM Trans. Netw. **9**, 6733 (2001).
[13] K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen, Phys. Rev. Lett. **90**, 148701 (2003).
[14] V. Batagelj and A. Mrvar, Connections **21**, 47 (1998).
[15] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabasi, Nature (London) **407**, 651 (2000).
[16] J. D. Noh and H. Rieger, Phys. Rev. Lett. **92**, 118701 (2004).
[17] L. da F. Costa and F. N. Silva, e-print cond-mat/0412761.
[18] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[19] P. G. Lind, M. C. Gonzalez, and H. J. Herrmann, Phys. Rev. E **72**, 056127 (2005).
[20] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, Proc. Natl. Acad. Sci. U.S.A. **98**, 4569 (2001).
[21] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, Phys. Rev. Lett. **94**, 028701 (2005).
[22] Website maintained by the NLANR Group at http://moat.nlanr.net/
[23] S. Valverde and R. V. Sole, Eur. Phys. J. B **38**, 245 (2004).
[24] A. Trusina, M. Rosvall, and K. Sneppen, Phys. Rev. Lett. **94**, 238701 (2005).
[25] M. Rosvall, P. Minnhagen, and K. Sneppen, Phys. Rev. E **71**, 066111 (2005).