Self-similar scale-free networks and disassortativity

Soon-Hyung Yook,* Filippo Radicchi, and Hildegard Meyer-Ortmanns†

SES, International University Bremen, P.O.Box 750561, D-28725 Bremen, Germany

(Received 8 July 2005; published 26 October 2005)

Self-similar networks with scale-free degree distribution have recently attracted much attention, since these apparently incompatible properties were reconciled in [C. Song, S. Havlin, and H. A. Makse, Nature **433**, 392 (2005)] by an appropriate box-counting method that enters the measurement of the fractal dimension. We study two genetic regulatory networks (*Saccharomyces cerevisiae* [N. M. Luscombe, M. M. Babu, H. Yu, M. Snyder, S. Teichmann, and M. Gerstein, Nature **431**, 308 (2004)] and *Escherichia coli* [http://www.ccg.unam.mx/Computational_Genomics/regulondb/DataSets/RegulonNetDataSets.html and http://www.gbf.de/SystemsBiology]) and show their self-similar and scale-free features, in extension to the datasets studied by [C. Song, S. Havlin, and H. A. Makse, Nature **433**, 392 (2005)]. Moreover, by a number of numerical results we support the conjecture that self-similar scale-free networks are not assortative. From our simulations so far these networks seem to be disassortative instead. We also find that the qualitative feature of disassortativity is scale-invariant under renormalization, but it appears as an intrinsic feature of the renormalization prescription, as even assortative networks become disassortative after a sufficient number of renormalization steps.

I. INTRODUCTION

Until very recently, the celebrated properties of a scale-free degree distribution seemed to be incompatible with self-similar features of networks, in which the number of boxes N_B of linear size ℓ_B scales with ℓ_B according to a power-law $N_B \sim \ell_B^{-d_B}$, with an exponent that is given by the fractal dimension d_B. Using the box-counting method, Song *et al.* showed that many scale-free (SF) networks observed in nature can have a fractal structure as well [1]. This result is striking, because the tiling and renormalization according to the linear size of the boxes, in which all pairs of nodes inside a box have mutual distance less than ℓ_B, appear to be physically relevant rather than being a formal procedure. Therefore, the essential quantity in the tiling is the linear size of the box, ℓ_B, defined by $\ell_B - 1$ being the maximal distance between the nodes of the box. In the renormalization procedure the boxes are contracted to the nodes of the renormalized network whose edges are the interconnecting edges between the boxes on the original network.

In this paper we study the genetic regulatory network of two well-known organisms, *Saccharomyces cerevisiae* [2] and *Escherichia coli* [3]. We first determine the degree-distribution $P(k)$, that is the probability for finding a node with degree k, i.e., with k edges, to read off the exponent γ according to $P(k) \sim k^{-\gamma}$ [4] in order to check that the networks are scale-free. The edges correspond to regulatory interactions between transcription factors and target genes (either between transcription factors and nontranscription factor targets, or between two transcription factors). Next we measure the ratio of the total number of boxes N_B of linear size

ℓ_B over the total number of nodes N in the network, that is N_B/N, for various box-sizes ℓ_B, to obtain the fractal dimension d_B from $N_B/N \sim \ell_B^{-d_B}$. After renormalizing the networks according to the procedure proposed in [1], we measure the scaling behavior of the degree k' according to $k' = s(\ell_B)k$, where k' stands for the degree of a node in the renormalized network, k is the largest degree inside the box that was contracted to one node with degree k' in the renormalization process, and $s(\ell_B)$ is assumed to scale like $s(\ell_B) \sim \ell_B^{-d_k}$ with a new exponent d_k. The invariance of γ under renormalization and the transformation behavior of the degree itself imply the relation [1]

$$\gamma = 1 + d_B/d_k \qquad (1)$$

between the exponents. Therefore, we check this relation by measuring d_B, d_k and comparing the values of γ from Eq. (1) with the measured γ from the degree distribution.

One of the important features of networks is their "degree" of assortativity. The notion of assortative mixing was known from epidemiology and ecology when it was introduced as a characteristic feature of generic networks by Newman [5]. Assortativity refers to correlations between properties of adjacent nodes. One particular property is the (in- or out-) degree of a node as the number of its incoming or outgoing links, respectively. Degree-degree correlations can be recorded as histograms; in order to facilitate the comparison between networks of different size, they can be also characterized by the Pearson coefficient. The Pearson coefficient is obtained from the connected degree-degree correlation function $\langle jk \rangle - \langle j \rangle \langle k \rangle$ after normalizing by its maximal value, which is achieved on a perfectly assortative network. Here, $\langle jk \rangle$ stands for the average of having vertex degrees j and k at the end of an arbitrary edge. The Pearson coefficient r takes values between $-1 \le r \le 1$, it is positive for assorta-

*Present address: Dept. of Physics and Research Institute of Basic Sciences, Kyung Hee University, Seoul, 130–701, Korea.

†Electronic address: h.ortmanns@iu-bremen.de

tive networks ($r=1$ for complete assortativity) and negative for disassortative ones. As it was shown in [5], the Pearson coefficient can be rewritten according to

$$r = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}, \quad (2)$$

where M denotes the total number of edges i, $i=1,\ldots,M$, and j_i, k_i denote the degrees of the two vertices at the ends of edge i. Equation (2) has actually the form in which we measured the Pearson coefficient for a number of self-similar networks. The results are presented below. The reason why this feature is of interest in the present context is its relation to the power-law or exponential behavior of $N_B(\ell_B)$. In particular, we are interested in the question whether disassortativity is scale-invariant on a qualitative level under renormalization according to the prescription proposed in [1], and why these properties go along. Disassortative features in protein interaction networks were found and explained by Maslov and Sneppen [6] on the level of interacting proteins and genetic regulatory interactions. According to their results links between highly connected nodes are systematically suppressed, while those between highly connected and low-connected pairs of proteins are favored. In this way there is little cross-talk between different functional modules of the cell and protection against intentional attacks, since the failure of one module is less likely to spread to another one. Also in immunological networks one speaks of lock- and key-interactions between molecular receptors and antigenic determinants [7]. In general, complementarity is essential for pattern recognition interactions, underlying biological and biochemical processes as well as for symbiotic species in ecological networks. Of course, it is not at all obvious or necessary that complementarity in "internal" (functional) properties should be manifest in topological features like the degree-degree correlations. Therefore, we study the relation between self-similarity and degree-assortativity in this paper.

II. MEASUREMENTS AND RESULTS

For the genetic regulatory networks *Saccharomyces cerevisiae* [2] and *Escherichia coli* [3] we observe a power-law behavior of N_B/N for $\ell_B > 3$ with $d_B = 5.1 \pm 0.3$ for *S. cerevisiae* and $d_B = 3.4 \pm 0.2$, respectively. The obtained degree-distributions are scale-free and satisfy a power-law with exponent $\gamma \simeq 2.6 \pm 0.2$ for *S. cerevisiae* and $\gamma \simeq 2.8 \pm 0.3$ for *E.coli*, cf. Fig. 1. The scaling relation (1) between the exponents d_B and d_k is also satisfied within the error bars for both networks. From (1) we obtain $\gamma = 2.8 \pm 0.3$ for *Saccharomyces cerevisiae* and $\gamma = 2.7 \pm 0.5$ for *Escherichia coli*.

In Table I we summarize the results also for some additional networks, for which we list their properties of self-similarity and disassortativity. If we confirm the property of self-similarity it means not only the scaling behavior of N_B/N according to a power-law, but also the numerical verification of the scaling relation of Eq. (1) and the invariance of γ under renormalization [1]. This is more conclusive, be-
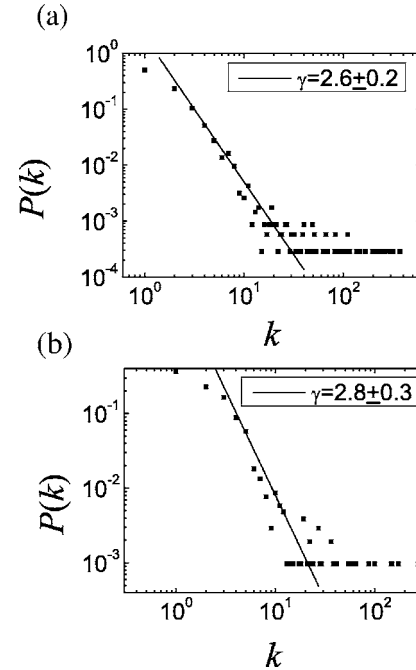

FIG. 1. Degree distribution $P(k)$ for (a) *S. cerevisiae* and (b) *E. coli* to read off γ.

cause it is sometimes difficult to disentangle exponential from power-law behavior of N_B/N for networks with a small diameter (for example, see Fig. 2 for the regulatory network of *S.cerevisiae* with an inset that shows the same data points on a log-linear scale instead of the log-log scale), whereas the scaling relation only holds for a power-law of the decay, it is easier to prove or disprove. A confirmation of (dis)assortativity refers to histograms with (negative) positive slope of next-neighbor degree-degree correlations and/or a (negative) positive Pearson coefficient, respectively. In most cases we measured the degree-degree correlations also between nodes at distance $d=2,3$, as indicated in the figures, where the distance is measured in units of edges.

The first four networks of Table I refer to the genetic regulatory networks of *S.cerevisia* and *E.coli*, the scientific collaboration network, and the internet on the autonomous systems level. For these networks the properties of columns 2 and 3 were examined by us, while for the last three networks (the biochemical pathway network of *E.coli*, the actor network and the world-wide-web), the self-similarity was es-

TABLE I. Networks from datasets characterized by properties of self-similarity and disassortativity.

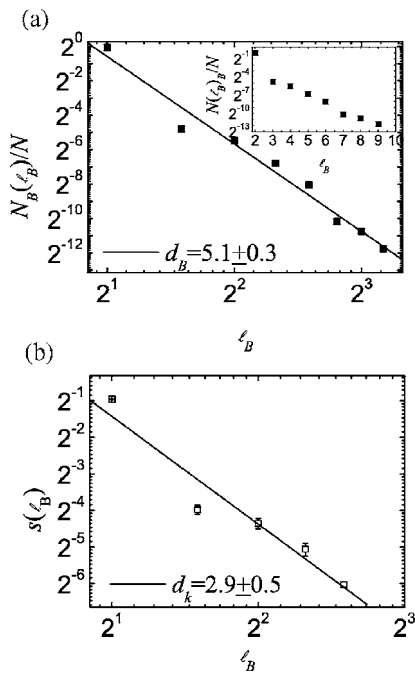| Network	Self-similarity	Disassortativity
genetic reg.*S.cerevisia* [2]	yes Fig. 2	yes Fig. 5(a)
genetic reg.*E.coli* [3]	yes Fig. 4	yes Fig. 5(b)
scient.collab. [8]	no Fig. 7(a)	no Fig. 7(b)
internet aut.sys. [8]	yes [10]	yes [10]
biochem.pathway *E.coli* [8]	yes [1]	yes [10]
actor [8]	yes [1]	(yes, $k>1000$) Fig. 3
www [8]	yes [1]	yes [10]

FIG. 2. *S.cerevisiae*: (a) Normalized number of boxes N_B as a function of linear box size ℓ_B to read off d_B. (b) Rescaling factor $s(\ell_B)$ as function of the box size to read off d_k.

tablished before [1], and we studied their property of disassortativity in addition. In particular the actor network deserves some further comments. The actor-network is self-similar [1], but its positive Pearson coefficient suggests that it is assortative, in contrast to all other self-similar networks we have studied so far. A closer look at its next neighbor-degree-degree correlation (Fig. 3) shows an assortative behavior for degrees up to the order of 1000, but slowly decays for larger degrees and becomes disassortative. The degree-degree correlation between nodes at distance larger than one is decreasing with degree k for all k.

Moreover, some comments are in order to the yeast-genetic regulatory network with 3456 nodes and 14 117 edges, (cf. Fig. 2). Since it has a diameter of 9, the largest ℓ_B value for the tiling is 10. Therefore, we have only 8 data points available for the fit. Each point corresponds to an
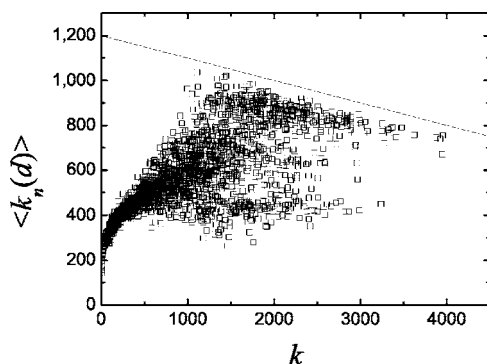

FIG. 3. Degree-degree correlation $\langle k_n(d)\rangle$ for $d=1$ against k for the actor network, showing initially assortative behavior for $k \leqslant 1000$.
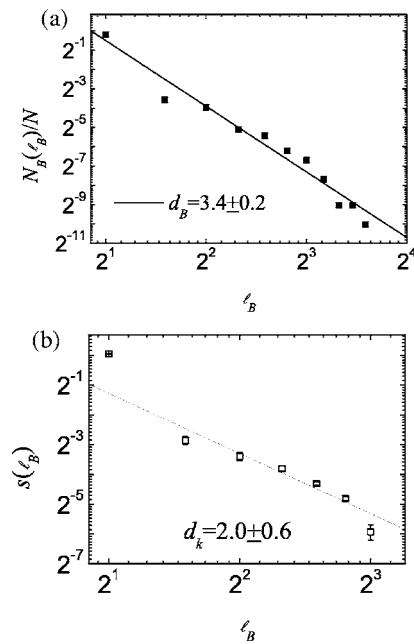

FIG. 4. Same as Fig. 2, but for *E.coli* to read off (a) d_B (b) d_k.

average over 100 tiling configurations. Different tiling configurations result from different starting seeds as well as the random selection of neighbors during the tiling process. The data point at $\ell_B=3$ in Fig. 2 lies clearly outside the fluctuations about the average over different tiling configurations, thus outside the error bars, which are at least two orders of magnitude smaller than the respective value of N_B/N, so that they are not visible on the scale of the figure. In Fig. 4, we find a similar behavior for *E.coli*. The deviation from the power-law behavior at $\ell_B=3$ goes along with an assortative degree-degree-correlation between nodes at distance $d=2$ as
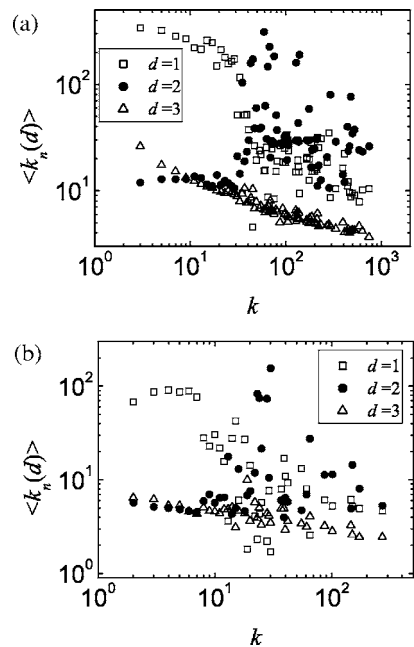

FIG. 5. Degree-degree correlation $\langle k_n(d)\rangle$ vs degree k for distances $d=1,2,3$, (a) *S.cerevisiae* and (b) *E.coli*.
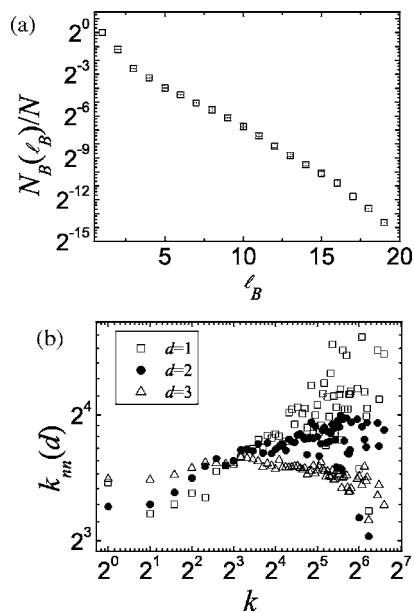
FIG. 6. Scientific collaboration network (a) exponential decay of $N_B/N(\ell_B)$ (b) degree-degree correlation $\langle k_n(d) \rangle$ for $d=1,2,3$ against k.

FIG. 7. Pearson coefficient r for several data sets, in particular the PIN-data taken from [9].

it is seen from Figs. 5(a) and 5(b), showing the degree-degree correlation of *S.cerevisiae* and *E.coli*, respectively, at distances $d=1,2,3$. The data in Fig. 5 explicitly show the disassortative behavior at $d=1$ and $d=3$ for both *S.cerevisiae* and *E.coli*. However, for $d=2$, we find that there is a certain value of $k,k=k^*$, at which $\langle k_n \rangle$ abruptly increases and slowly decreases for $k>k^*$. Here $k^* \sim 30$ for *S.cerevisiae* and $k^* \sim 10$ for *E.coli*. These mixed properties of assortativity and disassortativity seem to go along with the deviation from the power-law behavior of N_B/N. On a qualitative level, this is plausible if we focus on a hub that should be present in a scale-free network. In an assortative network (assortative say at distance d, for example, $d=2$), this hub is likely connected to another hub within the distance d. If this hub is chosen as a seed of a box in a tiling with linear box size $\ell_B>d$, we need much less boxes to cover the many nodes in the neighborhood of the hub than in a disassortative network.

In a network which is assortative not only for a certain range of k, but for all k and at distances $d \geq 1$, like the scientific collaboration network, N_B/N actually decays faster than power-like for all ℓ_B, as it is seen from the exponential fit of Fig. 6(a).
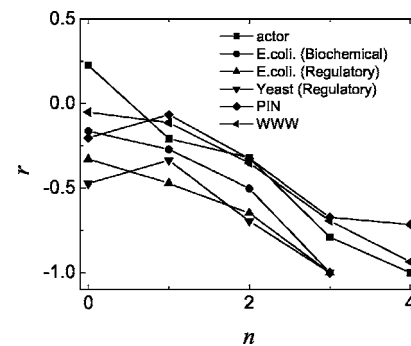
The scaling relation between the exponents γ, d_B, and d_k assumes the scale-invariance (under renormalization) of the degree distribution, that is the invariance of the exponent γ. Similarly, it is of interest how the disassortativity transforms under renormalization (as defined in [1]). As we see from Fig. 7, even networks like the scientific collaboration network (Fig. 6), which are originally assortative, transform to more and more disassortative ones under iterated renormalization. (The number of renormalization steps is determined by the size of the networks, in particular by its diameter. The final step is achieved when the reduced network consists of just one node.) Therefore the transformation behavior of disassortativity seems to be an effect of the renormalization procedure rather than an intrinsic self-similar property of the network. Similarly, we measured the transformation behavior of the clustering coefficient under renormalization of self-similar networks. As the data [10] show, it is an invariant property of scale-free networks, while it changes under renormalization for non-self-similar ones like the Barabási-Albert one [11].

To summarize, we find numerical evidence that self-similar scale-free networks are preferably disassortative in their degree-degree correlations. For biological networks this result may reflect the complementarity in interactions that is observed on various levels, as mentioned in the introduction, although it is still far from being obvious.

[1] C. Song, S. Havlin, and H. A. Makse, Nature (London) **433**, 392 (2005).

[2] N. M. Luscombe, M. M. Babu, H. Yu, M. Snyder, S. Teich-mann, and M. Gerstein, Nature (London) **431**, 308 (2004).

[3] http://www.ccg.unam.mx/Computational_Genomics/regulondb/DataSets/RegulonNetDataSets.html and http://www.gbf.de/SystemsBiology

[4] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[5] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002); M. E.

J. Newman and J. Park, Phys. Rev. E **68**, 036122 (2003).

[6] S. Maslov and K. Sneppen, Science **296**, 910 (2002).

[7] M. Copelli, R. M. Zorzenon dos Santos, and J. S. Sa Martins, cond-mat/0110350 (2001).

[8] www.nd.edu/networks

[9] P. Uetz *et al.*, Nature (London) **403**, 623 (2000).

[10] The additional data are available on the web http://imperia.iu-bremen.de/ses/physics/ortmanns/36753/index.shtml

[11] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).