Effect of degree correlations on the loop structure of scale-free networks

Ginestra Bianconi and Matteo Marsili
The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy
(Received 11 November 2005; published 28 June 2006)

In this paper we study the impact of degree correlations in the subgraph statistics of scale-free networks. In particular we consider loops, simple cases of network subgraphs which encode the redundancy of the paths passing through every two nodes of the network. We provide an understanding of the scaling of the clustering coefficient in modular networks in terms of the maximal eigenvector of the average adjacency matrix of the ensemble. Furthermore we show that correlations affect in a relevant way the average number of Hamiltonian paths in a three-core of real world networks. We prove our results in the two-vertex correlated hidden variable ensemble and we check the results with exact counting of small loops in real graphs.

DOI: 10.1103/PhysRevE.73.066127 PACS number(s): 89.75.Hc, 89.75.Da, 89.75.Fb

I. INTRODUCTION

The dynamics and the function of many complex systems strongly affect their network structure [1–4]. In fact both large-scale properties (like scale-free degree distribution [5]) and local properties (like recurrence of small motifs [6,7]) must be selected for widespread robustness requirements and specific preferential uses in real graphs. A large number of different networks [1–3], from the Internet to the protein interaction networks in a cell, share a scale-free degree distribution $P(k)\sim k^{-\gamma}$ with $\gamma<3$ and a high clustering coefficient with respect to random Erdös-Renyi graphs [8]. The scale-free degree distribution of a network affects the statistics of subgraphs present in it, showing that large-scale properties and local properties of scale-free networks are strongly related to each other. Special examples of subgraphs in networks are loops [9,10], paths that pass through each node in the loop only once. In random scale-free networks there are many small-size loops compared to random graphs and there can be a lack of Hamilton cycles (loops of length $L=N$) due to the fact that most of the large paths need to pass through hubs [10]. Along with other properties, many real scale-free networks also have degree correlations [11]. Degree correlations in real networks indicate that links are not randomly wired and that the probability that two nodes of degree k_i and k_j are linked deviates from the expected value $r_{i,j}=k_ik_j/(\langle k\rangle N)$. Consequently, correlated networks have at least one of the three following features (i) a k-dependent average connectivity $k_{nn}(k)$ of the first neighbors of a node with degree k [12,13]; (ii) a nontrivial dependence on the connectivity of the clustering coefficient $C=C(k)$ of nodes of degree k [14]; (iii) a cutoff that is larger than the structural cutoff $K\sim\sqrt{\langle k\rangle N}$. In particular many real scale-free networks show a power-law dependence on k both for $k_{nn}(k)$ and for $C(k)$, i.e., $k_{nn}(k)\sim k^{\alpha}$ and $C(k)\sim k^{-\delta}$ [12]. Correlations do affect the subgraph statistics as shown in the Internet [15] and in calculations based on the scaling of the clustering coefficient [7,16]. Every network can be represented in terms of its adjacency matrix (a) of elements $a_{i,j}=1,0$ depending if there is a link between node i and node j. From a formal point of view an ensemble of networks is given when a probability $\mathcal{P}(a)$ is assigned to each adjacency matrix (a) of $N\times N$ elements. In an uncorrelated and undirected network ensemble with given degree sequence $\{k_i\}$ all the links are independent. Consequently all the matrix elements $a_{i,j}$ with $i<j$ are independent and their average value in the ensemble can be written as $\langle a_{i,j}\rangle=r_{i,j}=\frac{k_ik_j}{\langle k\rangle N}$. A two-vertex correlated network is a network in which still the matrix elements $a_{i,j}$ with $i<j$ are independent but $\langle a_{i,j}\rangle=r_{i,j}\neq\frac{k_ik_j}{\langle k\rangle N}$. Networks with higher-order correlation instead would have nonindependent matrix elements which will favor some specific motifs in the network. In this paper we are going to provide an analytic calculation of the number of loops in two-vertex correlated scale-free networks. In the light of our results we are able to interpret the scaling of the clustering coefficient $C(k)$ in terms of the scaling of the maximal eigenvector (the eigenvector associated with the maximal eigenvalue) of the average adjacency matrix of the network ensemble. Moreover we show that the maximal eigenvalue and the corresponding eigenvector not only determine the number of triangles in the two-vertex correlated network ensemble, but also fix the number of small loops of length $3\ll L\ll N$. Finally, we are able to give a sufficient condition for the absence of Hamilton cycles in two-vertex correlated networks. This allows us to study a set of real graphs [the Internet at the autonomous system (AS) level and protein-protein interaction networks] [17] and show that, assuming they are specific instances of two-vertex correlated network ensembles, one can exclude the presence of Hamiltonian cycles in the three-core of these graphs. Our findings are in agreement for the Internet with what was found in Ref. [18] where a belief-propagation algorithm was applied to the measurement of the number of loops in real graphs. The absence of Hamiltonian cycles in a three-core of a network is an unexpected result since regular random graphs with connectivity $c\geqslant 3$ are Hamiltonian [8,19]. We note here that the average number $\langle\mathcal{N}_L\rangle$ of loops of size L in a two-vertex correlated network can possibly be dominated by a very large number of loops occurring in very rare networks [18]. Nevertheless, preliminary results indicate that in uncorrelated scale-free networks with $\gamma<3$ the ratio $\frac{\langle\mathcal{N}_L^2\rangle}{\langle\mathcal{N}_L\rangle^2}$ is bounded at least for small loops and for Hamiltonian cycles. We expect that similar arguments could also be extended to scale-free correlated networks.

The paper is organized as follows: In Sec. I we give an intuition of the results found for small loops in a two-vertex correlated network ensemble by considering the problem of

exact counting of loops in generic networks; in Sec. III we introduce the hidden variable ensemble and calculate the number of small loops and Hamiltonian cycles in two-vertex correlated hidden variable ensembles; in Sec. IV we compare the results with real networks; and finally we give the conclusions in Sec. V.

II. COUNTING SMALL LOOPS IN REAL NETWORKS

In this section we would like to provide some intuitive arguments to show that in scale-free networks the maximal eigenvalue of the adjacency matrix and the corresponding eigenvector are responsible for the number of small loops present in it. The adjacency matrix $((a))$ of a simple network of size N is the $N\times N$ matrix of elements $a_{i,j}=1,0$ indicating the existence $(a_{i,j}=1)$ or not $(a_{i,j}=0)$ of a link between node i and node j. The total number of closed paths of length L passing though a node i is given by the matrix element $(a^L)_{i,i}$. The loops $\mathcal{N}_L^{(i)}$ of size L passing through a node i are given by

$$\mathcal{N}_L^{(i)} = (a^L)_{i,i} - (\text{corrections}), \tag{1}$$

where these corrections account for closed paths which intersect themselves at least once and which must be subtracted from the term $(a^L)_{i,i}$ in order to consider only loops. If by λ_n we indicate the eigenvalues and by $\mathbf{u}^n$ the eigenvectors of the adjacency matrix (a) we find [9]

$$\mathcal{N}_L^{(i)} \sim \sum_n \lambda_n^L [u_i^{(n)} u_i^{(n)} - O(u_i^{(n)4})]. \tag{2}$$

For small L, the correction terms can be neglected if the spectrum of the graphs $\{\lambda\}$ contains one large eigenvalue $\lambda_0=\Lambda_0$ and if the associated normalized eigenvectors satisfy $0<u_i^{(n)}\ll 1\,\forall\, i$, as is the case in most scale-free networks. If these conditions are satisfied the sum over n in (2) is dominated by the term $n=0$ and consequently the number of loops of length L passing through the node i is given by

$$\mathcal{N}_L^{(i)} \sim \Lambda_0^L u_i^0 u_i^0, \tag{3}$$

while the total number of loops of size L is given by

$$\mathcal{N}_L = \frac{1}{2L}\sum_i \mathcal{N}_L^{(i)} \sim \frac{\Lambda_0^L}{2L}, \tag{4}$$

where the factor $2L$ accounts for the multiplicity of nodes a single loop passes through and the two possible directions of each loop. Thus we found by intuitive arguments that the total number of small loops of size L of scale-free networks will scale like Λ_0^L while the number of small loops passing through a node is proportional to the square of the maximal eigenvector associated with Λ_0. These arguments apply for the exact counting of small loops in real networks. In a random graph ensemble the adjacency matrix is a random variable which has average values of the elements $\langle a_{i,j}\rangle = r_{i,j}$ and we need to evaluate the average number of loops $\langle \mathcal{N}_L\rangle$ instead of $\mathcal{N}_L$. The results we will prove in the following sections are an extension of the expressions (4) and (3) to a two-vertex correlated hidden variable network ensemble.

III. AVERAGE NUMBER OF LOOPS IN CORRELATED HIDDEN VARIABLE ENSEMBLE

To model a general two-vertex correlated network in the following we will consider networks that are generated within the hidden variable model [21,20]. The prescription of Ref. [20] to generate a class of scale-free networks with exponent γ is the following, (1) assign to each node i of the graph a hidden continuous variable q_i distributed according to a $\rho(q)$ distribution. Then (2) each pair of nodes with hidden variables q,q' are linked with probability $r(q,q')$. When the hidden variable distribution is scale-free $\rho(q)=\rho_0 q^{-\gamma}$ for $q\in[m,Q]$ and $r(q,q')=qq'/(\langle q\rangle N)$, we obtain a random uncorrelated scale-free network. In this specific case a structural cutoff is needed to keep the linking probability smaller than 1, i.e., $Q^2/(\langle q\rangle N)<1$. This cutoff scales differently with the system size N depending on the value of γ: $Q\sim N^{1/(\gamma-1)}$ for $\gamma>3$, $Q\sim N^{1/2}$ for $\gamma\in(2,3)$, and $Q\sim N^{1/\gamma}$ for $\gamma\in(1,2)$. On the contrary, to generate a correlated scale-free network with natural cutoff $N^{1/(\gamma-1)}$ and $\gamma>2$ in the literature different *Ansätze* have been proposed [21,20]. In order to present general results on the average number of loops in the hidden variable ensemble for any type of linking probabilities $r(q,q')$ we consider an ordered set of distinct nodes $\{i_1,\ldots,i_n,\ldots,i_L\}$. With each such kind of set it is possible to associate a loop in the network in which subsequent nodes are linked with each other. For each choice of the nodes $\{i_1,\ldots,i_L\}$ with hidden variables $\{q_{i_1},\ldots,q_{i_L}\}$ the probability that they are connected in a loop is

$$r(q_{i_1},q_{i_2})r(q_{i_2},q_{i_3})\cdots r(q_{i_L},q_{i_1}) = \prod_n r(q_{i_n},q_{i_{mod(n+1,L)}}) \tag{5}$$

and for each loop of the network there are $2L$ ordered sets $\{i_1,\ldots,i_L\}$ which describe it corresponding to cyclic permutations of the indices and to their order inversion. The average number of loops of size L in the graph is given by the number of ways we can choose an ordered set of L nodes $\{i_1,\ldots,i_L\}$ multiplied by the probability that these nodes are connected in all distinguishable orderings and divided by $2L$. In order to proceed with the calculation, we lump together nodes with hidden variable $q_i\in[q,q+\Delta q)$, where Δq is a small interval of q. In each interval of q there are $N_q \simeq NP(q)\Delta q$ nodes of the network. For each choice of the L nodes, let n_q with $\Sigma_q n_q=L$ be the number of nodes in the loop with $q_{i_n}\in[q,q+\Delta q)$. The ways we can choose them within the N_q nodes of the network is given by the binomial $N_q!/[n_q!(N_q-n_q)!]$. Moreover let $n_{q,q'}$ indicate the nodes of a hidden variable q of the loop linked with a subsequent node of hidden variable q' in the fixed direction of the loop. We note that the way to choose $\{n_{q,q'}\}$ is given by the multinomial $n_q!/\Pi_{q'}n_{q,q'}!$ and that the partition $\{n_{q,q'}\}$ must satisfy the conditions $\Sigma_{q'}n_{q,q'}=n_q$ and $\Sigma_q n_{q,q'}=n_{q'}$. Finally the number of ways in which one can permute the L nodes keeping $n_{q,q'}$ constant is given by $\Pi_q n_q!$. Considering all this and that the probability Eq. (5) that the selected nodes are connected in the chosen order can be written as $\Pi_{q,q'}r(q,q')^{n_{q,q'}}$, we get the following expression for the average number of loops $\langle \mathcal{N}_L\rangle$ of size L:

$$\langle \mathcal{N}_L \rangle = \frac{1}{2L} \sum_{\{n_q\}}{}' \prod_q \frac{N_q!}{n_q!(N_q - n_q)!} \prod_q n_q! \sum_{\{n_{q,q'}\}}{}' \frac{n_q!}{\prod_{q'} n_{q,q'}!} \times \prod_{q,q'} r(q,q')^{n_{q,q'}} \tag{6}$$

where the sums $\Sigma'_{\{n_q\}}, \Sigma'_{\{n_{q,q'}\}}$ are extended over all $\{n_q\}$ and $\{n_{q,q'}\}$ such that $\Sigma_q n_q = L$, $\Sigma_{q'} n_{q,q'} = n_q$, and $\Sigma_q n_{q,q'} = n_{q'}$, and the factor $2L$ accounts for the multiplicity in which we count each loop. Introducing the constraints $\Sigma_q n_q = L$ and $\Sigma_q n_{q,q'} = n_{q'}$ by explicit δ functions, using their integral representation we find

$$\langle \mathcal{N}_L \rangle = \frac{1}{2L} \int_{-\infty}^{\infty} dx \sum_{\{n_q\}} e^{Lx} \prod_q \frac{N_q!}{n_q!(N_q - n_q)!} \prod_q n_q! e^{-x n_q} \times \int_{-\infty}^{\infty} \mathcal{D}x_q \prod_q e^{n_q x_q} \sum_{\{n_{q,q'}\}} \frac{n_q!}{\prod_{q'} n_{q,q'}!} \prod_{q,q'} r(q,q')^{n_{q,q'}} \times e^{-x_{q'} n_{q,q'}},$$

where $\mathcal{D}x_q$ indicates $\Pi_q dx_q$, and the sum over $\{n_{q,q'}\}$ is performed over all $\{n_{q,q'}\}$ such that $\Sigma_{q'} n_{q,q'} = n_q$. Consequently, performing the multinomial summations over $\{n_{q,q'}\}$ we get the following expressions:

$$\begin{aligned} \langle \mathcal{N}_L \rangle &= \frac{1}{2L} \int_{-\infty}^{\infty} dx\, e^{Lx} \sum_{\{n_q\}} \prod_q \frac{N_q!}{n_q!(N_q - n_q)!} e^{-x n_q} n_q! \\ &\quad \times \int_{-\infty}^{\infty} \mathcal{D}x_q \prod_q e^{n_q x_q} \left(\sum_{q'} r(q,q') e^{-x_{q'}} \right)^{n_q} \\ &= \frac{1}{2L} \int_{-\infty}^{\infty} dx\, e^{Lx} \sum_{\{n_q\}} \prod_q \frac{N_q!}{n_q!(N_q - n_q)!} \\ &\quad \times e^{-x n_q} n_q! \int_{-\infty}^{\infty} \mathcal{D}x_q e^{Q g(\{x_q\})} \end{aligned} \tag{7}$$

with

$$g(\{x_q\}) = \frac{1}{Q} \sum_q n_q \left[x_q + \ln \left(\sum_{q'} r(q,q') e^{-x_{q'}} \right) \right]. \tag{8}$$

Notice that in Eq. (7) one can safely take the limit $\Delta q \to 0$ and that the average over the $P(q)$ distribution is taken assuming that we focus on the limit $N \to \infty$. In what follows, we will evaluate Eq. (7) in different ranges of L in the limit $N \to \infty$. Assuming $L \gg 1$ we evaluate the integral over the variables $\{x_q\}$ by the saddle point equation finding

$$n_q = e^{-x_q} \sum_{q'} n_{q'} \frac{r(q',q)}{\sum_{q''} r(q',q'') e^{-x_{q''}}}. \tag{9}$$

If we indicate by $S_{q'}$ the sum $S_{q'} = \Sigma_{\bar{q}} r(q', \bar{q}) e^{-x_{\bar{q}}}$, we can cast the solution in the following form,

$$e^{-x_q} = n_q \frac{1}{\sum_{q'} n_{q'} r(q,q')/S'_q}. \tag{10}$$

This provides the self-consistent equation for $\{S_q\}$

$$S_q = \sum_{q'} n_{q'} \frac{r(q,q')}{\sum_{q''} n_{q''} r(q',q'')/S_{q''}}. \tag{11}$$

It is easy to check that $\{S_q\}$ satisfying the equation

$$S_q = \sum_{q'} n_{q'} \frac{r(q,q')}{S_{q'}} \tag{12}$$

is a solution of Eq. (11). Inserting a delta function $\delta(S_q - \Sigma_{q'} n_{q'} r(q,q')/S_{q'})$ and assuming that the Jacobian of this transformation is 1, i.e., assuming

$$S_q^2 \gg r(q,q'), \tag{13}$$

and using the Stirling approximation for the factorial n_q, the integrals over x_q calculated at the saddle point take the values $S_q^{2n_q} e^{-n_q \ln(n_q) + n_q}$ and the average number of loops of size L can be expressed as the following:

$$\langle \mathcal{N}_L \rangle = \int_{-\infty}^{\infty} dx\, e^{L(x-1)} \int \mathcal{D}S_q \int \mathcal{D}w_q \sum_{n_q}{}' \prod_q \frac{N_q!}{n_q!(N_q - n_q)!} \times (e^{-x} S_q^2)^{n_q} \exp \left[\sum_q w_q \left(S_q - \sum_{q'} n_{q'} \frac{r(q,q')}{S_{q'}} \right) \right]. \tag{14}$$

Finally, performing the summation over $\{n_q\}$ we get

$$\langle \mathcal{N}_L \rangle = \frac{1}{2L} \int_{-\infty}^{\infty} dx\, e^{L(x-1)} \int \mathcal{D}S_q \int \mathcal{D}w_q \prod_q \exp \left(N \left\{ \ln \left[1 + e^{-x} S_q^2 \exp \left(-N \sum_{q'} w_{q'} r(q,q')/S_q \right) \right] \right\} + N \sum_q w_q S_q \right), \tag{15}$$

where $\langle\ \rangle_q$ indicates the average over the distribution of the hidden variables N_q. In the limit $N > L \gg 1$ we evaluate the saddle point equations, finding

$$S_q = N\left\langle \frac{r(q,q')S_{q'}e^{-x}\exp\left(-\sum_{q''} r(q',q'')w_{q''}/S_{q'}\right)}{1+S_{q'}^2e^{-x}\exp\left(-\sum_{q''} r(q',q'')w_{q''}/S_{q'}\right)}\right\rangle_{q'},$$

$$w_q = -P(q) \times \frac{\left(2S_q+\sum_{q'} r(q,q')w_{q'}\right)e^{-x}\exp\left(-\sum_{q'} r(q,q')w_{q'}/S_q\right)}{1+S_q^2e^{-x}\exp\left(-\sum_{q'} r(q,q')w_{q'}/S_q\right)},$$

$$\ell = \left\langle \frac{S_q^2e^{-x}\exp\left(-\sum_{q'} r(q,q')w_{q'}/S_q\right)}{1+S_q^2e^{-x}\exp\left(-\sum_{q'} r(q,q')w_{q'}/S_q\right)}\right\rangle_q \quad (16)$$

with $\ell = L/N$. In order to solve these saddle point equations we make the *Ansätze*

$$N\sum_{q'} r(q,q')w_{q'} = \nu S_q. \quad (17)$$

With this assumption we can rewrite the saddle point equations (16) as

$$S_q = N\left\langle \frac{r(q,q')S_{q'}e^{-x-\nu}}{1+S_{q'}^2e^{-x-\nu}}\right\rangle_{q'},$$

$$w_q = -(2+\nu)P(q)\frac{S_qe^{-x-\nu}}{1+S_q^2e^{-x-\nu}},$$

$$\ell = \left\langle \frac{S_q^2e^{-x-\nu}}{1+S_q^2e^{-x-\nu}}\right\rangle_q, \quad (18)$$

which can be solved, and define the value of ν, $\nu=-1$.

A. The uncorrelated case

In the uncorrelated case, when $r(q,q')=\frac{qq'}{\langle q\rangle N}$ we found $S_q=q\sqrt{\frac{\ell}{\langle q\rangle}}$ which satisfies hypothesis (13). The results found in this limit are the same as the ones found in [10].

B. Small loops

The limit of small loop size is the limit $x \gg 1$. In this limit the saddle point equations (18) reduce to

$$S_q = \sum_{q'} N_{q'}r(q,q')S_{q'}e^{-x+1},$$

$$w_q = -P(q)r(q,q')S_qe^{-x+1},$$

$$\ell = \langle S_q^2\rangle_q e^{-x+1}. \quad (19)$$

The first equation indicates that S_q is the eigenvector of the average adjacency matrix $N_{q'}r(q,q')$ with eigenvalue $\Lambda = e^{x-1}$, the second equation defines the linear relation between w_q and S_q; and the third equation fixes the normalization constant for the eigenvector S_q. In this limit the average number of loops of size L is given by

$$\mathcal{N}_L \sim \frac{1}{2L}(\Lambda)^L \quad (20)$$

where Λ is the maximal eigenvalue of the average adjacency matrix $N_{q'}r(q,q')$, with the results valid until

$$\ell \ll \frac{\langle S_q^4\rangle}{\langle S_q^2\rangle^2}, \quad (21)$$

where S_q is the eigenvector of matrix $NP(q')r(q,q')$ corresponding to the maximal eigenvalue $\Lambda \gg \max S_q^2$. We observe that the vector $S_i = S_{q_i}$ with $i=1,\ldots,N$ is the eigenvector of the matrix $r_{i,j}=r(q_i,q_j)$. In other words $\{S_i\}$ is the eigenvector of the average adjacency matrix of the networks in the ensemble $\langle a_{i,j}\rangle = r_{i,j}$. This result provides the extension of the arguments of Sec. I, Eq. (4), to the two-vertex correlated network ensemble.

C. Small loops passing though a given node

From expression (15) one can also derive the number of small loops passing through a given node. One can easily show that

$$\mathcal{N}_L(q) \sim \frac{1}{2L}S_q^2\Lambda_{\{q\}}^{L-1} \quad (22)$$

where S_q is the maximal eigenvector of the matrix $N_{q'}r(q,q')$ normalized in such a way that $\langle S_q^2\rangle = \ell\Lambda$. This provides the extension of the arguments of Sec. I, Eq. (3), to a two-vertex correlated network ensemble.

D. Hamiltonian cycles

The Hamiltonian cycles of a graph are loops of size $L=N$. From Eq. (14) we find that when $L=N$ the expected number of Hamiltonian cycles goes to zero exponentially with N if

$$2\langle \ln(S_q)\rangle < 1 \quad (23)$$

with S_q satisfying

$$S_q = \sum_{q'} r(q,q')\frac{N_{q'}}{S_{q'}}. \quad (24)$$

Consequently, in the thermodynamic limit, since

$$P(\mathcal{N}_L > 0) \leqslant \langle \mathcal{N}_L\rangle, \quad (25)$$

(23) is a sufficient condition for excluding the presence of Hamiltonian cycles in the network.

IV. COMPARISON WITH REAL DATA

To test our calculation on real graphs and forecast some results regarding the existence or not of Hamiltonian cycles we have to assume that the real networks under study are a particular instance of a two-vertex correlated hidden variable network ensemble. Since the average connectivity $\bar{k}(q)$ of a node depends only on its hidden variable the minimal assumption one can make to fit real networks with the hidden

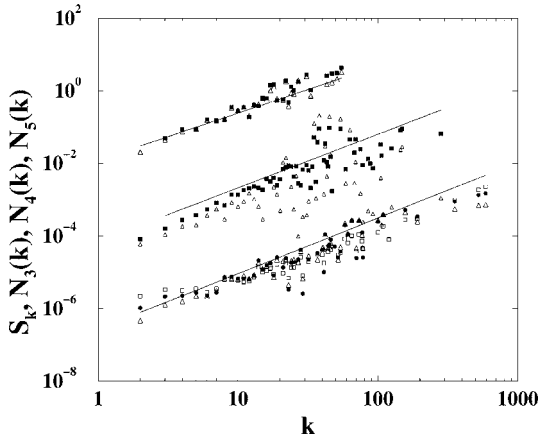

FIG. 1. Normalized number of triangles (empty triangles), quadrilaterals (filled squares), and pentagons (filled circles) passing through nodes of connectivity k. Data are shown for the Internet at the autonomous system level in November 1997 (bottom), in the *S. cerevisiae* protein interaction network (center), and in the *H. pylori* protein interaction network (top) [17]. The solid lines indicate the predictions $\mathcal{N}_L(k) \propto S_k^2$ where S_k is the maximal eigenvector of the correlation matrix $N_{k'} r(k,k')$. The data are shifted to improve the readability of the graph.

variable model is that the average degree is a one-to-one map to the hidden variable q. In this assumption maximum likelihood considerations force us to assume that each real graph is a random realization of a two-vertex correlated network with $q_i = k_i$ and $r(q=k, q=k') = \frac{N_{k,k'}}{\langle k \rangle N N_k N_{k'}}$ where $N_{k,k'}$ are the total number of links between nodes of degree k and k' and N_k and $N_{k'}$ are the numbers of nodes with degree k and k'.

This results give a very interesting interpretation of the dependence of the clustering coefficient on the connectivity k, i.e., $C(k) \sim \frac{1}{k(k-1)} \Lambda_{\{k\}}^2 S_k^2$ where S_k is the eigenvector associated with the maximal eigenvalue Λ of the matrix $N_{q'} r(q,q')$, in agreement with the intuitive arguments of Sec. I. In Fig. 1 we found that the data sets of the Internet and the protein interaction networks [17] have a loop structure well described by this two-vertex approximation. Moreover, one can predict if in the three-core of the considered graph there are no Hamiltonian cycles by evaluating if the condition (23) is satisfied, i.e., if

$$2\left\langle \ln(S_q) - \frac{1}{N}\langle \ln(p) \rangle \right\rangle < 1 \quad \text{with } S_q = \sum_{q'} r(q,q') \frac{N(q')}{S'_q},$$

where $\ln(p)/N = \langle \ln[1-(1+q+q^2/2)e^{-q}] \rangle$ corrects for the probability that the network in the ensemble contains nodes of connectivity $k<3$ as described in [10]. In particular one can compare the value of $2\langle \ln S_q \rangle$ calculated by solving (26) with $r(q=k, q'=k')$ extracted from the data $\left[r(q=k, q'=k') = \frac{N_{k,k'}}{\langle k \rangle N N_k N_{k'}} \right]$ with the value of $2\langle \ln S_q \rangle$ in the simplest example of a correlated ensemble, i.e., the static network ensemble [22] defined with $r(q=k, q'=k') = 1 - \exp\left[-\frac{kk'}{\langle k \rangle N} \right]$. We found as reported in Table I that the real degree correlations are such that the presence of Hamiltonian cycles in the three-core of the network is very unlikely.

TABLE I. Value of $2\langle \ln(S_q) \rangle - \langle \ln(p) \rangle / N$ with S_q satisfying Eq. (24) assuming as the maximum likelihood assumption that all the $q_i = k_i$ on the nodes of the three-core of Internet graphs, on various dates and on graphs of protein interactions [8]. We compare the value of $2\langle \ln(S_q) \rangle - \langle \ln(p) \rangle / N$ calculated with the two-vertex correlation assumption on real graphs or simply assuming the minimal assumption $r(q=k, q'=k') = 1 - e^{-kk'/\langle k \rangle N}$, i.e., $2\langle \ln(S_q^R) \rangle - \langle \ln(p^R) \rangle / N$. We observe that real correlations are essential to predict the absence of Hamiltonian cycles in these graphs.

	Network	$2\langle \ln(S_k) \rangle - \langle \ln(p) \rangle / N$	$2\langle \ln(S_k^R) \rangle - \langle \ln(p^R) \rangle / N$
AS	11–97	−4.73	2.98
	4–98	−5.22	3.06
	7–98	−5.35	3.03
	10–98	−5.56	3.01
	1–99	−5.74	3.07
	4–99	−6.06	3.09
	7–99	−6.28	3.07
	10–99	−6.55	3.06
	1–00	−6.75	3.07
	4–00	−7.20	3.01
	7–00	−7.30	3.03
	10–00	−7.46	3.01
	1–01	−7.428	3.01
	3–01	−7.73	3.00
DIP	*S. cerevisiase*	−6.46	3.99
	H. pylori	−4.5	3.8
	C. elegans	−0.66	2.89

V. CONCLUSIONS

In conclusion we have evaluated the number of loops of any size in two-vertex correlated networks. The results can be applied to real graphs, finding very good agreement of the predicted scaling of the clustering coefficient $C(k)$ with the square of the maximal eigenvector S_k of the matrix $N_{k'} r(k,k')$, i.e., $C(k) \sim S_k^2$. Moreover we can have a condition for predicting the absence of Hamiltonian cycles for the three-core of Internet and protein-protein interaction data. The results indicate that degree correlations strongly affect the loop frequency. Further study would consider how important are fluctuations of the number of loops around this average and would consider the frequency of other subgraphs in correlated scale-free networks.

ACKNOWLEDGMENTS

The work was supported by EVERGROW, integrated project No. 1935 in the complex systems initiative of the Future and Emerging Technologies directorate of the IST Priority, EU Sixth Framework and by EU Grant No. HPRN-CT-2002-00319, q.

[1] R. Albert and A.-L. Barabeási, Rev. Mod. Phys. **74**, 47 (2002).
[2] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University Press, Oxford, 2003).
[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet* (Cambridge University Press, Cambridge, U.K., 2004).
[5] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[6] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).
[7] A. Vazquez, R. Dobrin, D. Sergi, J.-P. Eckmann, Z. N. Oltvai, and A.-L. Barabási, Proc. Natl. Acad. Sci. U.S.A. **101**, 17940 (2004).
[8] S. Janson, T. Luczak, and A. Rucinski, *Random Graphs* (John Wiley & Sons, New York, 2000).
[9] G. Bianconi and A. Capocci, Phys. Rev. Lett. **90**, 078701 (2003).
[10] G. Bianconi and M. Marsili, J. Stat. Mech.: Theory Exp. P06005 (2005).
[11] J. Berg and M. Lassig, Phys. Rev. Lett. **89**, 228701 (2002).
[12] R. Pastor-Satorras, A. Vazquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
[13] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).
[14] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabasi, Science **297**, 1551 (2002).
[15] G. Bianconi, G. Caldarelli, and A. Capocci, Phys. Rev. E **71**, 066116 (2005).
[16] S. N. Soffer and A. Vazquez, Phys. Rev. E **71**, 057101 (2005).
[17] The Internet data sets are the ones collected by University of Oregon Route Views project and the protein-protein interaction data sets are those listed in the database of interacting proteins (DIP) database.
[18] E. Marinari, R. Monasson, and G. Semerjian, Europhys. Lett. **73**, 8 (2006).
[19] E. Marinari and R. Monasson, J. Stat. Mech.: Theory Exp. P09004 (2004).
[20] G Caldarelli, A. Capocci, P. De Los Rios, and M. A. Muñoz, Phys. Rev. Lett. **89**, 258702 (2002).
[21] M. Boguña and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003).
[22] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).