Financial Econometrics Research Centre

WORKING PAPERS SERIES

WP06-16

Effects of degree correlations on the loop structure of scale free networks

Ginestra Bianconi and Matteo Marsili

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

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.

PACS numbers: : 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, 2, 3, 4]. In fact both large-scale properties (like the 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, 2, 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 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 $r_{i,j} = k_i k_j/(\langle k\rangle N)$. 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 non trivial 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 $C(k) \sim k^{-\delta}$. 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 $i < j$ of are independent and their average value in the ensemble can be written as $\langle a_{i,j}\rangle = r_{i,j} = \frac{k_i k_j}{\langle k\rangle N}$. A two-vertex correlated network is a network in which still the matrix elements $a_{i,j}$ with $i < j$ are independent but $\langle a_{i,j}\rangle = r_{i,j} \neq \frac{k_i k_j}{\langle k\rangle N}$. Networks with higher order correlations instead would have non independent 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 they 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-DIP) [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 $c \geq 3$ 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. II we give an intuition of the results found for small loops in 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 we calculate the number of small loops and Hamiltonian cycles in two vertices 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]

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

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 pass 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 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 [20, 21]. 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 one, 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 ansatz have been proposed [20, 21]. 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 $\{i_1, \dots, i_n, \dots, i_L\}$. 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 $\{i_1, \dots, i_L\}$, with hidden variables $\{q_{i_1}, \dots, q_{i_L}\}$ 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 $\{i_1, \dots, i_L\}$ 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 $\{i_1, \dots, i_L\}$ 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 $\sum_q n_q = L$ 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 $n_q!/\prod_{q'} n_{q,q'}!$ and that the partition $\{n_{q,q'}\}$ must satisfy the conditions $\sum_{q'} n_{q,q'} = n_q$ and $\sum_q n_{q,q'} = n_{q'}$. Finally the number of ways in which one can permute the L nodes keeping $n_{q,q'}$ constant is given by $\prod_q n_q!$. 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,

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

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

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

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

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

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 $S_{q'} = \sum_{\bar{q}} r(q', \bar{q}) e^{-x_{\bar{q}}}$, 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

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

is a solution of the Eq. (11). Inserting a delta function $\delta\left(S_q - \sum_{q'} n_{q'} r(q,q')/S_{q'}\right)$ 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:

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

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

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

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

$$\begin{aligned}
S_q &= N \left\langle \frac{r(q,q') S_{q'} e^{-x} \exp\left(-\sum_{q''} r(q',q'') w_{q''}/S_{q'}\right)}{1 + S_{q'}^2 e^{-x} \exp\left(-\sum_{q''} r(q',q'') w_{q''}/S_{q'}\right)} \right\rangle_{q'} \\
w_q &= -P(q) \frac{(2S_q + \sum_{q'} r(q,q') w_{q'}) e^{-x} \exp\left(-\sum_{q'} r(q,q') w_{q'}/S_q\right)}{1 + S_q^2 e^{-x} \exp\left(-\sum_{q'} r(q,q') w_{q'}/S_q\right)} \\
\ell &= \left\langle \frac{S_q^2 e^{-x} \exp\left(-\sum_{q'} r(q,q') w_{q'}/S_q\right)}{1 + S_q^2 e^{-x} \exp\left(-\sum_{q'} r(q,q') w_{q'}/S_q\right)} \right\rangle_q .
\end{aligned} \tag{15}$$

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

$$N \sum_{q'} r(q,q') w_{q'} = \nu S_q. \tag{16}$$

With this assumption we can rewrite the saddle point equations (15) as

$$\begin{aligned}
S_q &= N \left\langle \frac{r(q,q') S_{q'} e^{-x-\nu}}{1 + S_{q'}^2 e^{-x-\nu}} \right\rangle_{q'}, \\
w_q &= -(2+\nu) P(q) \frac{S_q e^{-x-\nu}}{1 + S_q^2 e^{-x-\nu}}, \\
\ell &= \left\langle \frac{S_q^2 e^{-x-\nu}}{\cdots} \right.
\end{aligned} \tag{17}$$

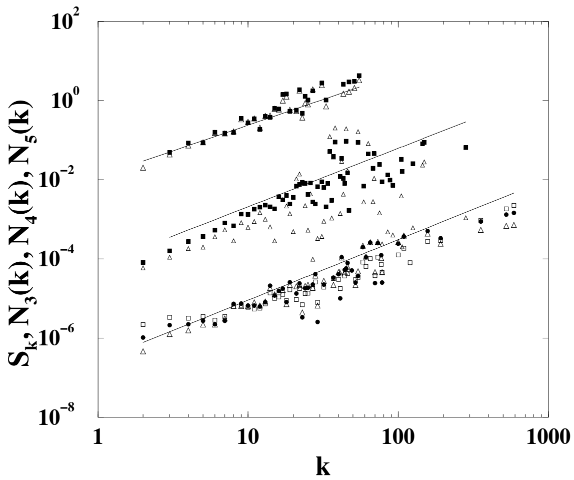

FIG. 1: Normalized number of triangles (empty triangles), quadrilaterals (filled squares) and pentagons (filled circles), passing through nodes of connectivity k. Data are shown for the Internet at the Autonomous System Level in November 1997 (bottom), in the s. cerevisiae protein interaction database (center) and in the h. pylori protein interaction (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.

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 (17) reduce to

$$\begin{aligned} S_q &= \sum_{q'} N_{q'} r(q, q') S_{q'} e^{-x+1} \\ w_q &= -P(q) r(q, q') S_q e^{-x+1} \\ \ell &= \langle S_q^2 \rangle_q e^{-x+1}. \end{aligned} \quad (18)$$

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

$$\mathcal{N}_L \sim \frac{1}{2L} (\Lambda)^L \quad (19)$$

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

TABLE I: In the table we report the value of $2\langle \ln(S_q) \rangle - \langle \log(p) \rangle / N$ with S_q satisfying Eq. (23) assuming as the maximum likelihood assumption that all the $q_i = k_i$ on the nodes of the three-core of the Internet graphs and on the graphs of protein interactions [17]. 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.

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

$$\ell \ll \frac{\langle S_q^4 \rangle}{\langle S_q^2 \rangle^2}, \quad (20)$$

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 $i = 1, \dots, N$ 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

$$\mathcal{N}_L(q) \sim \frac{1}{2L} S_q^2 \Lambda_{\{q\}}^{L-1} \quad (21)$$

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 section 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

$$2\langle \ln(S_q) \rangle < 1 \tag{22}$$

with S_q satisfying

$$S_q = \sum_{q'} r(q,q') N_{q'} / S'_q. \tag{23}$$

Consequently, in the thermodynamic limit, since

$$P(\mathcal{N}_L > 0) \leq \langle \mathcal{N}_L \rangle \,, \tag{24}$$

(22) if 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 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 networks 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. $C(k) \sim \frac{1}{k(k-1)} \Lambda^2_{\{k\}} S^2_k$ 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. Moreover, one can predict if in the three-core of the considered graph there are no Hamiltonian cycles by evaluating if the condition (22) is satisfied, i.e. if

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

where the $\ln(p)/N = \langle \ln[1 - (1 + q + q^2/2)e^{-q}] \rangle$ 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 (25) with $r(q = k, q' = k')$ extracted from the data $[r(q = k, q' = k') = \frac{N_{k,k'}}{\langle k \rangle N N_k N_{k'}}]$ 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 $r(q = k, q' = k') = 1 - \exp[-\frac{kk'}{\langle k \rangle N}]$. 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.

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. $C(k) \sim S^2_k$. 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.

VI. ACKNOWLEGMENTS

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 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 Review* **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 *PNAS* **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, JSTAT P06005 (2005).
[11] J. Berg and M. Lassig *Phys. Rev. Lett.* **89** (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. Barabási, *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**, (2005).
[17] The the Internet datasets are the one collecetd by University of Oregon Route Views project, NLANR and the protein-protein interaction datasets are one listed in DIP database.
[18] E. Marinari, R. Monasson and G. Semerjian, cond-mat/0507525 (2005).
[19] E. Marinari and R. Monasson, JSTAT 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.-L. Goh, B. Kahng and D. Kim *Phys. Rev. Lett.* **87**, 278701 (2001).

List of other working papers:

2006

1. Roman Kozhan, Multiple Priors and No-Transaction Region, WP06-24
2. Martin Ellison, Lucio Sarno and Jouko Vilmunen, Caution and Activism? Monetary Policy Strategies in an Open Economy, WP06-23
3. Matteo Marsili and Giacomo Raffaelli, Risk bubbles and market instability, WP06-22
4. Mark Salmon and Christoph Schleicher, Pricing Multivariate Currency Options with Copulas, WP06-21
5. Thomas Lux and Taisei Kaizoji, Forecasting Volatility and Volume in the Tokyo Stock Market: Long Memory, Fractality and Regime Switching, WP06-20
6. Thomas Lux, The Markov-Switching Multifractal Model of Asset Returns: GMM Estimation and Linear Forecasting of Volatility, WP06-19
7. Peter Heemeijer, Cars Hommes, Joep Sonnemans and Jan Tuinstra, Price Stability and Volatility in Markets with Positive and Negative Expectations Feedback: An Experimental Investigation, WP06-18
8. Giacomo Raffaelli and Matteo Marsili, Dynamic instability in a phenomenological model of correlated assets, WP06-17
9. Ginestra Bianconi and Matteo Marsili, Effects of degree correlations on the loop structure of scale free networks, WP06-16
10. Pietro Dindo and Jan Tuinstra, A Behavioral Model for Participation Games with Negative Feedback, WP06-15
11. Ceek Diks and Florian Wagener, A weak bifucation theory for discrete time stochastic dynamical systems, WP06-14
12. Markus Demary, Transaction Taxes, Traders' Behavior and Exchange Rate Risks, WP06-13
13. Andrea De Martino and Matteo Marsili, Statistical mechanics of socio-economic systems with heterogeneous agents, WP06-12
14. William Brock, Cars Hommes and Florian Wagener, More hedging instruments may destabilize markets, WP06-11
15. Ginwestra Bianconi and Roberto Mulet, On the flexibility of complex systems, WP06-10
16. Ginwestra Bianconi and Matteo Marsili, Effect of degree correlations on the loop structure of scale-free networks, WP06-09
17. Ginwestra Bianconi, Tobias Galla and Matteo Marsili, Effects of Tobin Taxes in Minority Game Markets, WP06-08
18. Ginwestra Bianconi, Andrea De Martino, Felipe Ferreira and Matteo Marsili, Multi-asset minority games, WP06-07
19. Ba Chu, John Knight and Stephen Satchell, Optimal Investment and Asymmetric Risk for a Large Portfolio: A Large Deviations Approach, WP06-06
20. Ba Chu and Soosung Hwang, The Asymptotic Properties of AR(1) Process with the Occasionally Changing AR Coefficient, WP06-05
21. Ba Chu and Soosung Hwang, An Asymptotics of Stationary and Nonstationary AR(1) Processes with Multiple Structural Breaks in Mean, WP06-04
22. Ba Chu, Optimal Long Term Investment in a Jump Diffusion Setting: A Large Deviation Approach, WP06-03
23. Mikhail Anufriev and Gulio Bottazzi, Price and Wealth Dynamics in a Speculative Market with Generic Procedurally Rational Traders, WP06-02
24. Simonae Alfarano, Thomas Lux and Florian Wagner, Empirical Validation of Stochastic Models of Interacting Agents: A "Maximally Skewed" Noise Trader Model?, WP06-01

2005

1. Shaun Bond and Soosung Hwang, Smoothing, Nonsynchronous Appraisal and Cross-Sectional Aggreagation in Real Estate Price Indices, WP05-17

2. Mark Salmon, Gordon Gemmill and Soosung Hwang, Performance Measurement with Loss Aversion, WP05-16
3. Philippe Curty and Matteo Marsili, Phase coexistence in a forecasting game, WP05-15
4. Matthew Hurd, Mark Salmon and Christoph Schleicher, Using Copulas to Construct Bivariate Foreign Exchange Distributions with an Application to the Sterling Exchange Rate Index (Revised), WP05-14
5. Lucio Sarno, Daniel Thornton and Giorgio Valente, The Empirical Failure of the Expectations Hypothesis of the Term Structure of Bond Yields, WP05-13
6. Lucio Sarno, Ashoka Mody and Mark Taylor, A Cross-Country Financial Accelorator: Evidence from North America and Europe, WP05-12
7. Lucio Sarno, Towards a Solution to the Puzzles in Exchange Rate Economics: Where Do We Stand?, WP05-11
8. James Hodder and Jens Carsten Jackwerth, Incentive Contracts and Hedge Fund Management, WP05-10
9. James Hodder and Jens Carsten Jackwerth, Employee Stock Options: Much More Valuable Than You Thought, WP05-09
10. Gordon Gemmill, Soosung Hwang and Mark Salmon, Performance Measurement with Loss Aversion, WP05-08
11. George Constantinides, Jens Carsten Jackwerth and Stylianos Perrakis, Mispricing of S&P 500 Index Options, WP05-07
12. Elisa Luciano and Wim Schoutens, A Multivariate Jump-Driven Financial Asset Model, WP05-06
13. Cees Diks and Florian Wagener, Equivalence and bifurcations of finite order stochastic processes, WP05-05
14. Devraj Basu and Alexander Stremme, CAY Revisited: Can Optimal Scaling Resurrect the (C)CAPM?, WP05-04
15. Ginwestra Bianconi and Matteo Marsili, Emergence of large cliques in random scale-free networks, WP05-03
16. Simone Alfarano, Thomas Lux and Friedrich Wagner, Time-Variation of Higher Moments in a Financial Market with Heterogeneous Agents: An Analytical Approach, WP05-02
17. Abhay Abhayankar, Devraj Basu and Alexander Stremme, Portfolio Efficiency and Discount Factor Bounds with Conditioning Information: A Unified Approach, WP05-01

2004

1. Xiaohong Chen, Yanqin Fan and Andrew Patton, Simple Tests for Models of Dependence Between Multiple Financial Time Series, with Applications to U.S. Equity Returns and Exchange Rates, WP04-19
2. Valentina Corradi and Walter Distaso, Testing for One-Factor Models versus Stochastic Volatility Models, WP04-18
3. Valentina Corradi and Walter Distaso, Estimating and Testing Sochastic Volatility Models using Realized Measures, WP04-17
4. Valentina Corradi and Norman Swanson, Predictive Density Accuracy Tests, WP04-16
5. Roel Oomen, Properties of Bias Corrected Realized Variance Under Alternative Sampling Schemes, WP04-15
6. Roel Oomen, Properties of Realized Variance for a Pure Jump Process: Calendar Time Sampling versus Business Time Sampling, WP04-14
7. Richard Clarida, Lucio Sarno, Mark Taylor and Giorgio Valente, The Role of Asymmetries and Regime Shifts in the Term Structure of Interest Rates, WP04-13
8. Lucio Sarno, Daniel Thornton and Giorgio Valente, Federal Funds Rate Prediction, WP04-12
9. Lucio Sarno and Giorgio Valente, Modeling and Forecasting Stock Returns: Exploiting the Futures Market, Regime Shifts and International Spillovers, WP04-11
10. Lucio Sarno and Giorgio Valente, Empirical Exchange Rate Models and Currency Risk: Some Evidence from Density Forecasts, WP04-10
11. Ilias Tsiakas, Periodic Stochastic Volatility and Fat Tails, WP04-09
12. Ilias Tsiakas, Is Seasonal Heteroscedasticity Real? An International Perspective, WP04-08
13. Damin Challet, Andrea De Martino, Matteo Marsili and Isaac Castillo, Minority games with finite score memory, WP04-07
14. Basel Awartani, Valentina Corradi and Walter Distaso, Testing and Modelling Market Microstructure Effects with an Application to the Dow Jones Industrial Average, WP04-06

15. Andrew Patton and Allan Timmermann, Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity, WP04-05
16. Andrew Patton, Modelling Asymmetric Exchange Rate Dependence, WP04-04
17. Alessio Sancetta, Decoupling and Convergence to Independence with Applications to Functional Limit Theorems, WP04-03
18. Alessio Sancetta, Copula Based Monte Carlo Integration in Financial Problems, WP04-02
19. Abhay Abhayankar, Lucio Sarno and Giorgio Valente, Exchange Rates and Fundamentals: Evidence on the Economic Value of Predictability, WP04-01

2002

1. Paolo Zaffaroni, Gaussian inference on Certain Long-Range Dependent Volatility Models, WP02-12
2. Paolo Zaffaroni, Aggregation and Memory of Models of Changing Volatility, WP02-11
3. Jerry Coakley, Ana-Maria Fuertes and Andrew Wood, Reinterpreting the Real Exchange Rate - Yield Diffential Nexus, WP02-10
4. Gordon Gemmill and Dylan Thomas , Noise Training, Costly Arbitrage and Asset Prices: evidence from closed-end funds, WP02-09
5. Gordon Gemmill, Testing Merton's Model for Credit Spreads on Zero-Coupon Bonds, WP02-08
6. George Christodoulakis and Steve Satchell, On th Evolution of Global Style Factors in the MSCI Universe of Assets, WP02-07
7. George Christodoulakis, Sharp Style Analysis in the MSCI Sector Portfolios: A Monte Caro Integration Approach, WP02-06
8. George Christodoulakis, Generating Composite Volatility Forecasts with Random Factor Betas, WP02-05
9. Claudia Riveiro and Nick Webber, Valuing Path Dependent Options in the Variance-Gamma Model by Monte Carlo with a Gamma Bridge, WP02-04
10. Christian Pedersen and Soosung Hwang, On Empirical Risk Measurement with Asymmetric Returns Data, WP02-03
11. Roy Batchelor and Ismail Orgakcioglu, Event-related GARCH: the impact of stock dividends in Turkey, WP02-02
12. George Albanis and Roy Batchelor, Combining Heterogeneous Classifiers for Stock Selection, WP02-01

2001

1. Soosung Hwang and Steve Satchell , GARCH Model with Cross-sectional Volatility; GARCHX Models, WP01-16
2. Soosung Hwang and Steve Satchell, Tracking Error: Ex-Ante versus Ex-Post Measures, WP01-15
3. Soosung Hwang and Steve Satchell, The Asset Allocation Decision in a Loss Aversion World, WP01-14
4. Soosung Hwang and Mark Salmon, An Analysis of Performance Measures Using Copulae, WP01-13
5. Soosung Hwang and Mark Salmon, A New Measure of Herding and Empirical Evidence, WP01-12
6. Richard Lewin and Steve Satchell, The Derivation of New Model of Equity Duration, WP01-11
7. Massimiliano Marcellino and Mark Salmon, Robust Decision Theory and the Lucas Critique, WP01-10
8. Jerry Coakley, Ana-Maria Fuertes and Maria-Teresa Perez, Numerical Issues in Threshold Autoregressive Modelling of Time Series, WP01-09
9. Jerry Coakley, Ana-Maria Fuertes and Ron Smith, Small Sample Properties of Panel Time-series Estimators with I(1) Errors, WP01-08
10. Jerry Coakley and Ana-Maria Fuertes, The Felsdtein-Horioka Puzzle is Not as Bad as You Think, WP01-07
11. Jerry Coakley and Ana-Maria Fuertes, Rethinking the Forward Premium Puzzle in a Non-linear Framework, WP01-06
12. George Christodoulakis, Co-Volatility and Correlation Clustering: A Multivariate Correlated ARCH Framework, WP01-05

13. Frank Critchley, Paul Marriott and Mark Salmon, On Preferred Point Geometry in Statistics, WP01-04
14. Eric Bouyé and Nicolas Gaussel and Mark Salmon, Investigating Dynamic Dependence Using Copulae, WP01-03
15. Eric Bouyé, Multivariate Extremes at Work for Portfolio Risk Measurement, WP01-02
16. Erick Bouyé, Vado Durrleman, Ashkan Nikeghbali, Gael Riboulet and Thierry Roncalli, Copulas: an Open Field for Risk Management, WP01-01

2000

1. Soosung Hwang and Steve Satchell , Valuing Information Using Utility Functions, WP00-06
2. Soosung Hwang, Properties of Cross-sectional Volatility, WP00-05
3. Soosung Hwang and Steve Satchell, Calculating the Miss-specification in Beta from Using a Proxy for the Market Portfolio, WP00-04
4. Laun Middleton and Stephen Satchell, Deriving the APT when the Number of Factors is Unknown, WP00-03
5. George A. Christodoulakis and Steve Satchell, Evolving Systems of Financial Returns: Auto-Regressive Conditional Beta, WP00-02
6. Christian S. Pedersen and Stephen Satchell, Evaluating the Performance of Nearest Neighbour Algorithms when Forecasting US Industry Returns, WP00-01

1999

1. Yin-Wong Cheung, Menzie Chinn and Ian Marsh, How do UK-Based Foreign Exchange Dealers Think Their Market Operates?, WP99-21
2. Soosung Hwang, John Knight and Stephen Satchell, Forecasting Volatility using LINEX Loss Functions, WP99-20
3. Soosung Hwang and Steve Satchell, Improved Testing for the Efficiency of Asset Pricing Theories in Linear Factor Models, WP99-19
4. Soosung Hwang and Stephen Satchell, The Disappearance of Style in the US Equity Market, WP99-18
5. Soosung Hwang and Stephen Satchell, Modelling Emerging Market Risk Premia Using Higher Moments, WP99-17
6. Soosung Hwang and Stephen Satchell, Market Risk and the Concept of Fundamental Volatility: Measuring Volatility Across Asset and Derivative Markets and Testing for the Impact of Derivatives Markets on Financial Markets, WP99-16
7. Soosung Hwang, The Effects of Systematic Sampling and Temporal Aggregation on Discrete Time Long Memory Processes and their Finite Sample Properties, WP99-15
8. Ronald MacDonald and Ian Marsh, Currency Spillovers and Tri-Polarity: a Simultaneous Model of the US Dollar, German Mark and Japanese Yen, WP99-14
9. Robert Hillman, Forecasting Inflation with a Non-linear Output Gap Model, WP99-13
10. Robert Hillman and Mark Salmon , From Market Micro-structure to Macro Fundamentals: is there Predictability in the Dollar-Deutsche Mark Exchange Rate?, WP99-12
11. Renzo Avesani, Giampiero Gallo and Mark Salmon, On the Evolution of Credibility and Flexible Exchange Rate Target Zones, WP99-11
12. Paul Marriott and Mark Salmon, An Introduction to Differential Geometry in Econometrics, WP99-10
13. Mark Dixon, Anthony Ledford and Paul Marriott, Finite Sample Inference for Extreme Value Distributions, WP99-09
14. Ian Marsh and David Power, A Panel-Based Investigation into the Relationship Between Stock Prices and Dividends, WP99-08
15. Ian Marsh, An Analysis of the Performance of European Foreign Exchange Forecasters, WP99-07
16. Frank Critchley, Paul Marriott and Mark Salmon, An Elementary Account of Amari's Expected Geometry, WP99-06
17. Demos Tambakis and Anne-Sophie Van Royen, Bootstrap Predictability of Daily Exchange Rates in ARMA Models, WP99-05
18. Christopher Neely and Paul Weller, Technical Analysis and Central Bank Intervention, WP99-04
19. Christopher Neely and Paul Weller, Predictability in International Asset Returns: A Re-examination, WP99-03

20. Christopher Neely and Paul Weller, Intraday Technical Trading in the Foreign Exchange Market, WP99-02
21. Anthony Hall, Soosung Hwang and Stephen Satchell, Using Bayesian Variable Selection Methods to Choose Style Factors in Global Stock Return Models, WP99-01

1998

1. Soosung Hwang and Stephen Satchell, Implied Volatility Forecasting: A Compaison of Different Procedures Including Fractionally Integrated Models with Applications to UK Equity Options, WP98-05
2. Roy Batchelor and David Peel, Rationality Testing under Asymmetric Loss, WP98-04
3. Roy Batchelor, Forecasting T-Bill Yields: Accuracy versus Profitability, WP98-03
4. Adam Kurpiel and Thierry Roncalli , Option Hedging with Stochastic Volatility, WP98-02
5. Adam Kurpiel and Thierry Roncalli, Hopscotch Methods for Two State Financial Models, WP98-01