Cycles and clustering in bipartite networks

Article *in* Physical Review E · December 2005

DOI: 10.1103/PhysRevE.72.056127 · Source: PubMed

3 authors, including:

Pedro G Lind
Oslo Metropolitan University
132 PUBLICATIONS **1,779** CITATIONS
SEE PROFILE

Marta C. Gonzalez
Massachusetts Institute of Technology
124 PUBLICATIONS **14,782** CITATIONS
SEE PROFILE

Cycles and clustering in bipartite networks

Pedro G. Lind,[1,2] Marta C. González,[1] and Hans J. Herrmann[1,3]

[1]*Institute for Computational Physics, Universität Stuttgart, Pfaffenwaldring 27, D-70569 Stuttgart, Germany*
[2]*Centro de Física Teórica e Computacional, Av. Prof. Gama Pinto 2, 1649-003 Lisbon, Portugal*
[3]*Departamento de Física, Universidade Federal do Ceará, 60451-970 Fortaleza, Brazil*

(Dated: February 2, 2008)

We investigate the clustering ability in bipartite networks where cycles of size three are absent and therefore the standard definition of clustering coefficient cannot be used. Instead, we use another coefficient given by the fraction of cycles with size four, showing that both coefficients yield the same clustering properties. The new coefficient is computed for two networks of sexual contacts, one monopartite and another bipartite. In both cases the clustering ability is similar. Furthermore, combining both clustering coefficients we deduce an expression for estimating cycles of larger size, which improves previous estimations and is suitable for either monopartite and multipartite networks.

PACS numbers: 89.75.Fb, 89.75.Hc, 89.65.-s
Keywords: Clustering Coefficient, Cycles, Bipartite Networks, Social Networks

I. INTRODUCTION

One important statistical tool to access the structure of complex networks arising in many systems [1, 2] is the clustering coefficient, introduced by Watts and Strogatz [3] to measure "the cliquishness of a typical neighborhood" in the network and given by the average fraction of neighbors which are interconnected with each other. This quantity has been used for instance to characterize small-world networks [3], to understand synchronization in scale-free networks of oscillators [4] and to characterize chemical reactions [5] and networks of social relationships [6, 7]. One pair of linked neighbors corresponds to a 'triangle', i.e. a cycle of three connections.

While triangles may be abundant in monopartite networks, they cannot be formed in bipartite networks [7, 8, 9], where two types of nodes exist and connections link only nodes of different type. Thus, the standard clustering coefficient is always zero. However, different bipartite networks have in general different cliquishnesses and clustering abilities [7], stemming for another coefficient which uncovers these topological differences among bipartite networks. Bipartite networks arise naturally in e.g. social networks [8] where the relationships (connections) depend on the gender of each person (node), and there are situations, such as in sexual contact networks [10], where one is interested in comparing clustering properties between monopartite and bipartite compositions.

In this paper, we study the cliquishness of either monopartite and bipartite networks, using both the standard clustering coefficient and an additional coefficient which gives the fraction of squares, i.e. cycles composed by four connections. As shown below, such a coefficient retains the fundamental properties usually ascribed to the standard clustering coefficient in regular, small-world and scale-free networks. As a specific application, two examples of networks of sexual contacts will be studied and compared, one being monopartite and another bipartite.

Furthermore, we will show that one can take triangles and squares as the basic units of larger cycles in any network, monopartite or multipartite. The frequency and distribution of larger cycles in networks have revealed its importance in recent research for instance to characterize local ordering in complex networks from which one is able to give insight on their hierarchical structure [11], to determine equilibrium properties of specific network models [12], to estimate the ergodicity of scale-free networks [13], to detect phase transitions in the topology of bosonic networks [14] and to help characterizing the Internet structure [15]. Since the computation of all cycles in arbitrarily large networks is unfeasible, one uses approximate numerical algorithms [13, 16, 17] or statistical estimations [18, 19]. Here, we go a step further and deduce an expression to estimate the number of cycles of larger size, using both clustering coefficients, which not only improves recent estimations [19] done for monopartite networks, but at the same time can be applied to bipartite networks and multipartite networks of higher order.

We start in Section II by introducing the expression which characterizes the cliquishness of bipartite networks, comparing it with the usual clustering coefficient. In Section III we apply both coefficients to real networks of sexual contacts and in Section IV we use them to estimate cycles of larger size and show how it is applied to bipartite networks. Conclusions are given in Section V.

II. TWO COMPLEMENTARY CLUSTERING COEFFICIENTS

The standard definition of clustering coefficient C_3 is the fraction between the number of triangles observed in one network out from the total number of possible triangles which may appear. For a node i with a number k_i of neighbors the total number of possible triangles is just the number of pairs of neighbors given by $k_i(k_i-1)/2$. Thus, the clustering coefficient $C_3(i)$ for node i is

$$C_3(i) = \frac{2t_i}{k_i(k_i-1)}. \qquad (1)$$

where t_i is the number of triangles observed, i.e. the number of connections among the k_i neighbors.

Similarly to $C_3(i)$, a cluster coefficient $C_4(i)$ with squares is the quotient between the number of squares and the total number of possible squares. For a given node i, the number of observed squares is given by the number of common neighbors among its neighbors, while the total number of possible squares is given by the sum over each pair of neighbors of the product between their degrees, after subtracting the common node i and an additional one if they are connected. Explicitly this clustering coefficient reads

$$C_4(i) = \frac{\sum_{m=1}^{k_i} \sum_{n=m+1}^{k_i} q_i(m,n)}{\sum_{m=1}^{k_i} \sum_{n=m+1}^{k_i} [a_i(m,n) + q_i(m,n)]}, \qquad (2)$$

where m and n label neighbors of node i, $q_i(m,n)$ are the number of common neighbors between m and n and $a_i(m,n) = (k_m - \eta_i(m,n))(k_n - \eta_i(m,n))$ with $\eta_i(m,n) = 1 + q_i(m,n) + \theta_{mn}$ and $\theta_{mn} = 1$ if neighbors m and n are connected with each other and 0 otherwise.

While $C_3(i)$ gives the probability that two neighbors of node i are connected with each other, $C_4(i)$ is the probability that two neighbors of node i share a common neighbor (different from i). Averaging $C_3(i)$ and $C_4(i)$ over the nodes yields two complementary clustering coefficients, $\langle C_3 \rangle$ and $\langle C_4 \rangle$, characterizing the contribution for the network cliquishness of the first and second neighbors respectively. For simplicity we write henceforth C_3 and C_4 for the averages of $C_3(i)$ and $C_4(i)$ respectively.

Figure 1 shows both clustering coefficients C_3 and C_4 in several topologies. In all cases C_3 and C_4 are plotted as dashed and solid lines respectively, and are averages over samples of 100 realizations. As an example of regular networks, we use networks with boundary conditions where each node has n neighbors symmetrically disposed. In particular, for $n = 2$ one obtains a chain of nodes. For these regular networks, Fig. 1a shows the dependence of the clustering coefficients on the fraction n/N of neighbors, with $N = 10^3$ the total number of nodes. As one sees $C_4 < C_3$ and for either small or large fractions of neighbors both coefficients increase abruptly with n. In the middle region C_3 is almost constant, while C_4 decreases slightly. Our simulations have shown that in regular networks the coefficients depend only on n/N, i.e. for any size of the regular network, similar plots are obtained.

Figure 1b shows the coefficients for small-world networks with $N = 10^3$ nodes, constructed from a regular network with $n = 4$ neighbors symmetrically disposed. The coefficients are computed as functions of the probability p to rewire short-range connections into long-range connections and they are normalized as usual [3] to the clustering coefficients $C^0_{3,4}$ of the underlying regular network. As one sees, C_4 yields approximately the same spectrum as the standard clustering coefficient C_3 being therefore able to define the same range of p for which small-world effects are observed. While here the small-world networks were constructed with rewiring of short-range connections into long-range ones, the same features are observed when using the construction procedure introduced in Ref. [20] where instead of rewiring one uses addition of long-range connections.

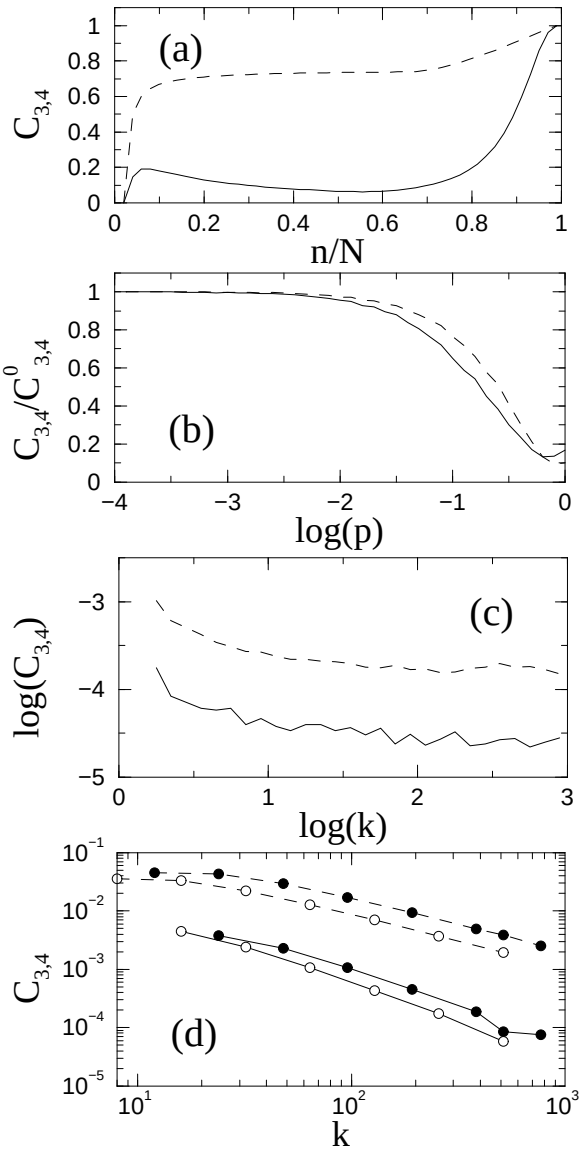

FIG. 1: Comparisons between the standard clustering coefficient C_3 in Eq. (1) (dashed line) and the clustering coefficient C_4 in Eq. (2) (solid line) for different network topologies: **(a)** in one regular network with n neighbors symmetrically placed ($N = 10^3$), **(b)** in small-world networks where long-range connections occur with probability p ($N = 10^3$ and $n = 4$) and **(c)** in random scale-free networks where the distribution of the clustering coefficients is plotted as a function of the number k of neighbors ($N = 10^5$ and $m = 2$). In all cases samples of 10^2 networks were used. The distributions $C_3(k)$ and $C_4(k)$ are also plotted for **(d)** Apollonian networks [21] with $N = 9844$ nodes (•) and pseudo-fractal networks [22] with $N = 9843$ nodes (○).

For random scale-free networks we plot in Fig. 1c the distribution of both coefficients as functions of the number k of neighbors, using networks with $N = 10^5$ nodes and by given initially $m = 2$ connections to each node. Here, one observes that $C_4(k)$ is almost constant as k increases, reproducing the same known feature as the standard $C_3(k)$ apart

a scaling factor: $C_4(k)/C_3(k)$ is approximately constant for any k. In Fig. 1d we plot the clustering distributions for two different deterministic scale-free networks recently studied, namely Apollonian networks [21], represented with bullets $\bullet$, and pseudo-fractal networks [22], represented with circles $\circ$. In both cases, the same power-law behavior already known for $C_3(k) \sim k^{-\alpha}$ in these hierarchical networks is also observed for the coefficient $C_4(k)$ with the same value of the exponent α.

In short, the results shown in Fig. 1 give evidence that C_4 is also a suitable coefficient to characterize the topological features in several complex networks commonly done with the standard clustering coefficient C_3. Furthermore, since C_4 counts squares instead of triangles, it is particularly suited for bipartite networks. Next, we will use this coefficient to compare different models for networks of sexual contacts, where both monopartite and bipartite networks arise naturally.

III. CYCLES AND CLUSTERING IN SEXUAL NETWORKS

In this Section we apply both coefficients C_3 and C_4 in Eqs. (1) and (2) to analyze two real networks of sexual contacts. One network is obtained from an empirical data set, composed solely by heterosexual contacts among $N = 82$ nodes, extracted at the Cadham Provincial Laboratory and is a 6-month block data [23] between November 1997 and May 1998. The other data set is the largest cluster with $N = 250$ nodes in the records of a contact tracing study [24], from 1985 to 1999, for HIV tests in Colorado Springs (USA), where most of the registered contacts were homosexual. Figure 2 sketches these two networks, where one can see that cycles of different sizes appear. While the network with only heterosexual contacts is clearly bipartite, the network with homosexual contacts is monopartite.

For the two networks in Fig. 2, Table I indicates the number T of triangles, the number Q of squares and the coefficients C_3 and C_4. As one sees, although the heterosexual network has less squares than the homosexual network due to its smaller size, C_4 is much larger. Another feature common for both neighbors is the average number of connections per node $L/N \sim 1$.

Recently, we introduced [10] a model to simulate the statistical features of these networks of sexual contacts. The model is a sort of a granular system with low density composed by N mobile particles representing persons and collisions between them representing their sexual contacts. The collisions representing sexual contacts are governed by dynamical rules which are carried out by means of an event-driven algorithm, and are based on two simple facts from sociological observations: (i) individuals with a larger number of partners are more likely to get new partners and (ii) sexual interactions do not determine the direction toward which each agent will be moving afterward. Therefore, we choose a collision rule where the absolute value of the velocity of each agent increases with the number k of sexual partners and the moving directions after collisions are randomly selected [10].

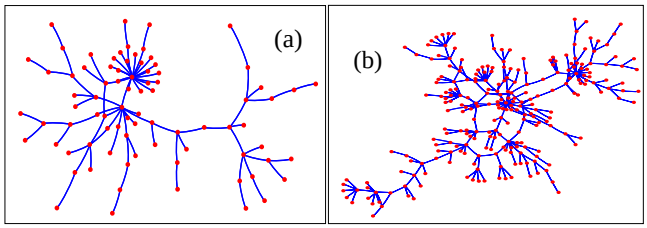

FIG. 2: Sketch of two real sexual contact networks having **(a)** only heterosexual contacts ($N = 82$ nodes and $L = 84$ connections) and **(b)** homosexual contacts ($N = 250$ nodes and $L = 266$ connections). While in the homosexual network triangles and squares appear, in the heterosexual network triangles are absent (see Table I).

	N	L	T	Q	$\langle C_3 \rangle$	$\langle C_4 \rangle$
Heterosexual (Fig. 2a)	82	84	0	2	0	0.00486
Homosexual (Fig. 2b)	250	266	11	6	0.02980	0.00192
Heterosexual (Agent Model)	82	83.63	0	1.45	0	0.01273
Homosexual (Agent Model)	250	287.03	8.23	10.52	0.02302	0.01224
Heterosexual (Scale-free)	82	162	0	159.72	0	0.12859
Homosexual (Scale-free)	250	498	45.28	256.79	0.08170	0.02787

TABLE I: Clustering coefficients and cycles in two real networks of sexual contacts (top), illustrated in Fig. 2, one where all contacts are heterosexual (Fig. 2a) and another with homosexual contacts (Fig. 2b). In each case one indicates the values of the number N of nodes, the number L of connections, the number T of triangles, the number Q of squares and both clustering coefficients C_3 and C_4 in Eqs. (1) and (2) respectively. The values of these quantities are also indicated for networks constructed with the agent model recently introduced [10] and for random scale-free networks.

Using the same number of nodes as in the real networks illustrated in Fig. 2 and considering two types of nodes for the heterosexual (bipartite) case, we obtain with the agent model similar results for L, T, Q, C_3 and C_4, as shown in Table I where values represent averages over samples of 100 realizations. Remarkably, for the bipartite case not only the number of connections and the number of squares are numerically the same, but also C_4 is of the same order of magnitude. Similar values of the topological quantities are also obtained for the monopartite case, with the exception of C_4. Despite this difference, the agent model gives values for the topological quantities of clustering and cycles much more closely to the real ones than in random scale-free networks, commonly used to reproduce such empirical data sets of sexual contacts [25]. In Table I we also show the values obtained for monopartite and bipartite scale-free networks whose degree distributions are as close as possible from the distributions of the real networks.

To compare more deeply scale-free networks and networks

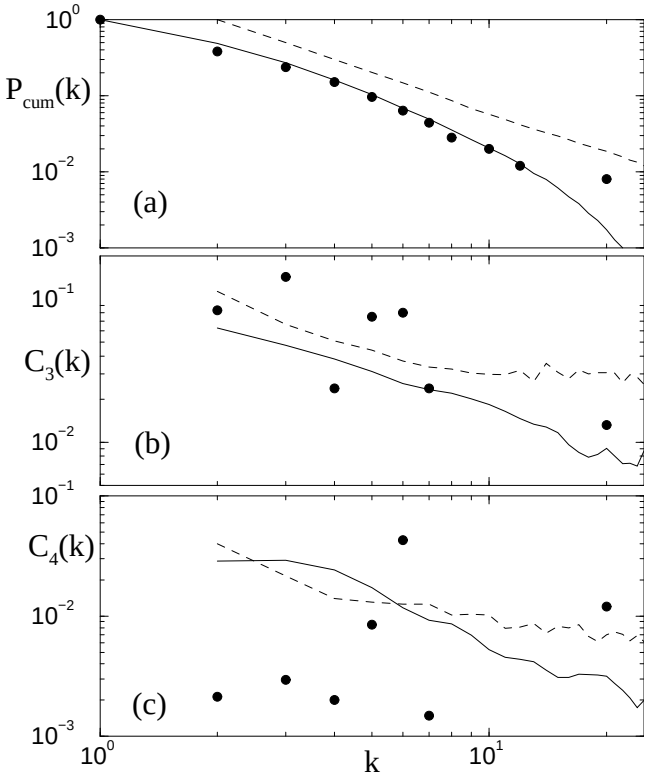

FIG. 3: Comparing topological features between networks obtained from the agent model (solid lines) and random scale-free networks (dashed lines), both used to reproduce one real monopartite network of sexual contacts (bullets): **(a)** cumulative degree distribution $P_{cum}(k)$, **(b)** standard clustering coefficient $C_3(k)$ in Eq. (1) and **(c)** clustering coefficient $C_4(k)$ in Eq. (2). Here $N = 250$ and samples of 100 realizations were used. For the scale-free network $m = 2$ which yields the best results for the coefficients (see text).

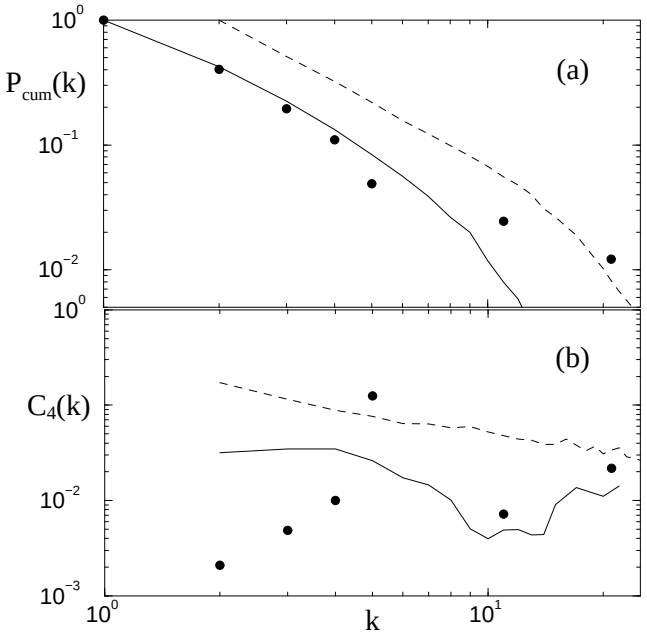

FIG. 4: Distributions for one real bipartite network of sexual contacts (bullets) compared with the one of networks obtained from the agent model (solid lines) and with random scale-free networks (dashed lines): **(a)** cumulative degree distribution $P_{cum}(k)$, **(b)** clustering coefficient $C_4(k)$ in Eq. (2). Here $C_3(k) = 0$ always, $N = 82$, samples of 100 realizations were used for both models and scale-free networks have $m = 2$.

obtained with the agent model we study also the distribution of the number k of sexual partners and the coefficients distributions as functions of k. In Fig. 3 we plot these distributions for the monopartite network of sexual contacts sketched in Fig. 2b, while in Fig. 4 we plot the distributions for the bipartite network (Fig. 2a). In both figures bullets indicate the distributions of the empirical data, while solid lines indicate the distributions of the networks obtained with the agent model and dashed lines indicate the distributions of scale-free networks, with a minimum number of connections $m = 2$, which gives the best fit of a scale-free distribution to the empirical data with non-zero clustering coefficient. For both models we impose the same size as the real network and take averages over a sample of 100 realizations.

As illustrated in Fig. 3a, the agent model reproduces accurately the cumulative degree distribution $P_{cum}(k)$ of the real monopartite network. For $m = 1$ the degree distribution of the scale-free network yields a better fit to the empirical data than for $m = 2$ but both clustering coefficients are zero for any degree k, which is not realistic as illustrated in Figs. 3b and 3c.

Figures 3b and 3c show the distribution of C_3 and C_4 respectively. Although the real network (bullets) is very small and therefore finite size effects appear, one may observe that C_3 is larger than C_4. Clearly, both models yield clustering coefficients of the same order of magnitude as the ones of the real networks, remaining the condition $C_3(k) > C_4(k)$ observed for the real network. For scale-free networks the clustering coefficients are slightly larger than the ones of the agent model.

Figures 4a and 4b show the cumulative degree distributions and the distribution of C_4 respectively, for the bipartite network of heterosexual contacts. Here, $C_3(k) = 0$ for all k (not shown). As one sees, the cumulative distribution for the agent model yields a better fit to the empirical distribution than the one of scale-free networks, as seen in Fig. 4a. Notice that the cumulative degree distribution for the scale-free network deviates from a power-law at large values of k, since we are using a bipartite graph which decrease the number of the most connected nodes. Furthermore, comparing Fig. 4b with Fig. 3c one clearly sees that in both cases, C_4 has approximately the same shape.

It is interesting to observe that, while both models reproduce at least qualitatively the coefficient distributions, in all cases the agent model fits more accurately the degree distribution of the empirical data. Furthermore, comparing C_4 between homosexual and heterosexual contacts one may rise the hypothesis that the cliquishness of both types of contacts is similar (see Sec. V below).

In the next Section we will present another application of the clustering coefficient C_4, showing how it can be used to account for cycles of larger size in any network, and in particular in bipartite networks.

IV. ESTIMATING THE NUMBER OF LARGE CYCLES WITH SQUARES AND TRIANGLES

Recent studies have attracted attention to the cycle structure of complex networks, since the presence of cycles has important effects for example on information propagation through the network [27] and on epidemic spreading behavior [28]. In order to avoid numerical algorithms for counting the number of cycles with arbitrary size which implies long computation times, an estimate of the fraction of cycles with different sizes was proposed [19], using the degree distribution $P(k)$ and the standard cluster coefficient distribution $C_3(k)$. However, this estimation yields a lower bound for the total number of cycles and cannot be applied to bipartite networks, as shown below. In this Section we show that by using both C_3 and C_4 one is able to improve that estimation, being suitable at the same time to either monopartite and bipartite networks.

The estimation in Ref. [19] considers the set of cycles with a central node, i.e. cycles with one node connected to all other nodes composing the cycle. Figure 5a illustrates one of such cycles, where the central node and each pair of its consecutive neighbors forms a triangle, in a total amount of four adjacent triangles. In such set of cycles, to estimate the number of cycles with size s one looks to the central node of each cycle which has a number, say k, of neighbors. The number of different possible cycles to occur is $n_0(s,k) = \binom{k}{s-1}\frac{(s-1)!}{2}$, since one has $\binom{k}{s-1}$ different groups of s nodes and in each one of these groups there are $(s-1)!/2$ different ways in ordering the s nodes into a cycle. The fraction of $n_0(s,k)$ of cycles which is expected to occur is $p_0(s,k) = C_3(k)^{s-2}$, since the probability of having one edge between two consecutive neighbors is $C_3(k)$ and one must have $s-2$ edges between the $s-1$ neighbors. Therefore, the number of cycles of size s is estimated as

$$N_s \;=\; N g_s \sum_{k=s-1}^{k_{max}} P(k) n_0(s,k) p_0(s,k), \qquad (3)$$

where $P(k)$ is the degree distribution and g_s is a factor which takes into account the number of repeated cycles.

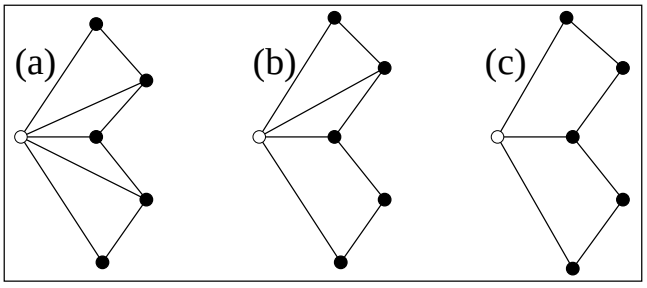

FIG. 5: Illustrative examples of cycles (size $s = 6$) where the most connected node ($\circ$) is connected to **(a)** all the other nodes composing the cycle, forming four adjacent triangles. In **(b)** the most connected node is connected to all other nodes except one, forming two triangles and one sub-cycle of size $s = 4$, while in **(c)** the same cycle $s = 6$ encloses two sub-cycles of size $s = 4$ and no triangles (see text).

The estimation in Eq. (3) is a lower bound for the total number of cycles since it considers only cycles with a central node. For instance, in Fig. 5b while cycles of size $s = 4$ can be estimated with Eq. (3), the cycle $s = 6$ cannot since it has no central node, and in Fig. 5c the above equation cannot estimate any cycle of any size. In fact, Fig. 5c illustrates the type of cycles appearing in bipartite networks, where no triangles are observed. For such cycles $C_3(k) = 0$ and therefore all terms in Eq. (3) vanish yielding a wrong estimation of the number of cycles.

To take into account cycles without central nodes (Figs. 5b and 5c), one must consider the clustering coefficient $C_4(k)$ defined in Eq. (2). One first considers the set of cycles of size s with one node ($\circ$) connected to all the others *except* one, as illustrated in Fig. 5b. In this case, since there are $s-2$ nodes connected to node $\circ$ one has $n_1(s,k) = \binom{k}{s-2}(s-2)!/2$ different possible cycles of size s, with k the number of neighbors of node $\circ$. The fraction of the $n_1(s,k)$ cycles which is expected to be observed is given by $p_1(s,k) = C_3(k)^{s-4}C_4(k)(1-C_3(k))$, since the probability of having $s-4$ connections among the $s-2$ connected nodes is $C_3(k)^{s-4}$, the probability that a pair of neighbors of node $\circ$ has to share a common neighbor (different from node $\circ$) is $C_4(k)$ and the probability that these same pair of neighbors have to be not connected is $(1-C_3(k))$. Writing an equation similar to Eq. (3), where instead of $n_0(s,k)$ and $p_0(s,k)$ one has $n_1(s,k)$ and $p_1(s,k)$ respectively and the sum starts at $s-2$ instead of $s-1$, one has an additional number N'_s of estimated cycles which are not considered in estimation (3). Notice that, since for N'_s one considers at least one sub-cycle of size $s = 4$, this additional estimation contributes only for the estimation of cycles with size $s \geq 4$. We call henceforth sub-cycle, a cycle which is enclosed in a larger cycle and which do not enclose itself any shorter cycle.

Still, the new estimation $N_s + N'_s$ is not suitable for bipartite networks, since it yields nonzero estimation only for $s = 4$. To improve the estimation further one must consider not only cycles composed by one single sub-cycle of size $s = 4$, as done in the previous paragraph, but also cycles with any number of sub-cycles of size $s = 4$. Figure 5c illustrates a cycle of size $s = 6$ composed by two sub-cycles of size 4. In general, following the same approach as previously, for cycles composed by q sub-cycles of size 4 one finds $n_q(s,k) = \frac{(s-q-1)!}{2}\binom{k}{s-q-1}$ possible cycles of size s looking from a node with k neighbors and a fraction $p_q(s,k) = C_3(k)^{s-2q-2}C_4(k)^q(1-C_3(k))^q$ of them which are expected to be observed. For $q = 0$ one considers cycles as the one illustrated in Fig. 5a, while for $q = 1$ and $q = 2$ one considers the set of cycles with one and two sub-cycles with size 4, as illustrated in Figs. 5b and 5c respectively. Summing

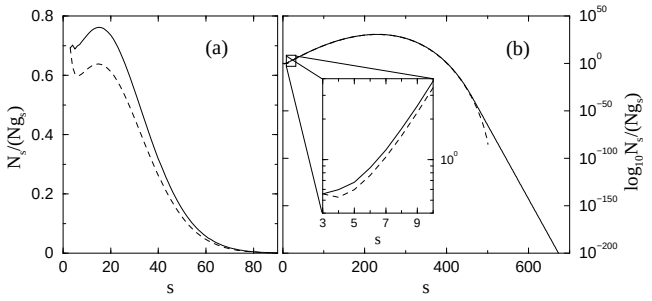

FIG. 6: Estimating the number of cycles using Eq. (3), dashed lines, and Eq. (4), solid lines. Here we impose a degree distribution $P(k) = P_0 k^{-\gamma}$ with $P_0 = 0.737$ and $\gamma = 2.5$, and coefficient distributions $C_{3,4}(k) = C_{3,4}^{(0)} k^{-\alpha}$ with **(a)** $C_3^{(0)} = 2$, $C_4^{(0)} = 0.33$, $\alpha = 0.9$ and **(b)** $C_3^{(0)} = 1$, $C_4^{(0)} = 0.17$, $\alpha = 1.1$. In all cases $k_{max} = 500$.

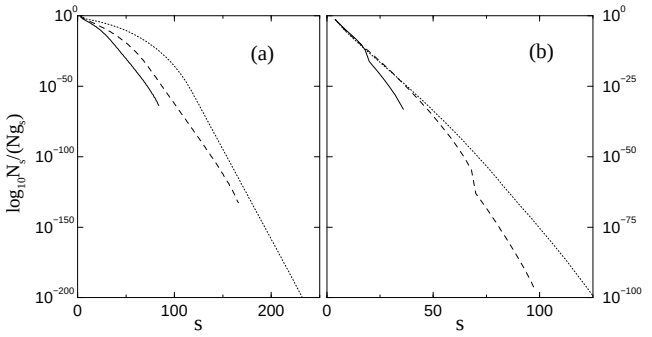

FIG. 7: Estimating the number of cycles for the agent model using Eq. (4) for $N = 1000$ (solid lines), $N = 5000$ (dashed lines) and $N = 10000$ (dotted lines) in **(a)** a monopartite network and in **(b)** a bipartite network, both obtained with the agent model.

up over k and q yields our final expression

$$N_s = N g_s \sum_{q=0}^{[s/2]-1} \sum_{k=s-q-1}^{k_{max}} P(k) n_q(s,k) p_q(s,k). \quad (4)$$

where $[x]$ denotes the integer part of x. In particular, the first term ($q = 0$) is the sum in Eq. (3). The upper limit $[s/2]-1$ of the first sum results from the fact that the exponent of $C_3(k)$ in $p_q(s,k)$ must be non-negative: $s-2q-2 \geq 0$. The estimation in Eq. (4) not only improves the estimated number computed from Eq. (3), but also enables the estimation of cycles up to a larger maximal size. In fact, since in the binomial coefficient $\binom{k}{s-1}$ of Eq. (3) one must have $s-1 \leq k \leq k_{max}$, one only estimates cycles of size up to $k_{max}+1$, while in Eq. (4) the maximal size is $2k_{max}$, as can be concluded using both conditions $s-2q-2 \geq 0$ and $s-q-1 \leq k_{max}$.

Figure 6 compares two cases treated in Ref. [19], both with a degree distribution $P(k) = P_0 k^{-\gamma}$ and coefficient distributions $C_3(k) = C_3^{(0)} k^{-\alpha}$, using one value of $\alpha < 1$ (Fig. 6a) and another one $\alpha > 1$ (Fig. 6b). Dashed lines indicate the estimation done with Eq. (3), while solid lines indicate the estimation done with Eq. (4). In both cases, the latter estimation is larger. For $\alpha < 1$ the difference between both estimations decreases with the size s of the cycle. For $\alpha > 1$ the difference between the estimations increases with s beyond a size $s^* \lesssim k_{max}$. Clearly, from Fig. 6b one sees that $k_{max}+1$ is the larger cycle size for which Eq. (3) can give an estimation, while for Eq. (4) the estimation proceeds up to $2k_{max}$ (partially shown). In both cases, the typical size for which N_s attains a maximum is numerically the same for both estimations, as expected. Moreover, for $\alpha > 1$ (Fig. 6b), beyond a size of the order of k_{max}, $N_s/(Ng_s)$ in Eq. (4) decreases exponentially with s, and not as a cutoff as observed for Eq. (3). In fact, the deviation of Eq. (3) from the exponential tail, is due to the fact that for very large cycle sizes ($s \sim k_{max}$) Eq. (3) can only consider very few terms in its sum.

Another advantage of the estimation in Eq. (4) is that it estimates cycles in bipartite networks. For bipartite network there are no connections between the neighbors, i.e. all subgraphs are similar to the one illustrated in Fig. 5c. Therefore all terms in Eq. (4) vanish except those for which the exponent of $C_3(k)$ is zero, i.e. for $s = 2(q+1)$. Consequently, since q is an integer, Eq. (4) shows clearly that in bipartite networks there are only cycles of even size, as already known [8]. Moreover, substituting $q = (s-2)/2$ in Eq. (4) yields a simple expression for the number of cycles in bipartite networks, namely

$$N_s^{\text{Bipart}} = N g_s \sum_{k=s/2}^{k_{max}} P(k) \frac{(s/2)!}{2} \binom{k}{s/2} C_4(k)^{s/2-1}. \quad (5)$$

Figure 7 shows distribution of the fraction $N_s/(Ng_s)$ of cycles as a function of their size s, for a monopartite network (Fig. 7a) and a bipartite network (Fig. 7b) composed by $N = 1000, 5000$ and 10000 nodes. This networks were generated from the agent model described in the previous section. Here, while monopartite networks show an exponential tail preceded by a region where the number of cycles is large, bipartite networks are composed by cycles whose number depend exponentially of their size. Furthermore one observes a clear transition for a characteristic size, which seems to scale with the network size.

It is important to notice that triangles and squares may appear in any multipartite network (except in bipartite ones, where triangles are absent). Therefore, the estimation described and studied in this Section can be applied not only to bipartite networks but to any multipartite network of any order.

V. DISCUSSION AND CONCLUSIONS

We introduced a clustering coefficient similar to the standard one, which instead of measuring the fraction of triangles in a network measures the fraction of squares, and showed that with this clustering coefficient it is also possible to characterize topological features in complex networks, usually done with the standard coefficient. We showed explicitly that the

range of values of the probability to acquire long-range connections in small-world networks and the typical clustering coefficient distributions of either random scale-free and hierarchical networks are approximately the same. In addition, we showed that this second clustering coefficient enables one to quantify the cliquishness in bipartite networks where triangles are absent. Thus, one should take triangles and squares simultaneously as the two basic cycle units in any network.

An application of both clustering coefficients was proposed, namely to estimate the number of cycles in any network, either monopartite or multipartite. Using a recent estimation which yields a lower bound of the number of cycles in monopartite network up to a size $s < k_{max} + 1$ where k_{max} is the maximum number of neighbors in the network, we deduce a more general expression which not only improves the previous estimation but is also suitable for bipartite networks and enables one to estimate cycles of size up to $2k_{max}$. Furthermore, in the particular case of bipartite networks our estimation yields as a natural consequence that only cycles of even size may appear.

We also studied a concrete example of two sexual networks, one where only heterosexual contacts occur (bipartite network) and another with homosexual contacts (monopartite). The results obtained with the two real networks were compared with the ones obtained with a scale-free network and with an agent model recently introduced. Our results emphasize that, in general, the agent model seems to be more suitable to reproduce networks of sexual contacts than the standard approach with random scale-free networks. Furthermore, our results for the clustering distribution of both real sexual networks gave some evidence that the clustering ability in sexual networks probably does not depend on the type of sexual contact (homosexual or heterosexual). To strengthen this hypothesis it is necessary to use larger networks of sexual contacts and apply the topological quantities here described. These and other questions are being analyzed in more detail and will be presented elsewhere.

Acknowledgments

MCG thanks Deutscher Akademischer Austausch Dienst (DAAD), Germany, and PGL thanks Fundação para a Ciência e a Tecnologia (FCT), Portugal, for financial support.

[1] B. Bollobás and O.M. Riordan, *Handbook of Graphs and Networks: From the Genome to the Internet* (Wiley-VCH, Weinheim, 2003).
[2] M.E.J. Newman, SIAM Review **45**(2), 167-256 (2003).
[3] D.J. Watts and S.H. Strogatz, Nature **393**, 440-442 (1998).
[4] P.N. McGraw and M. Menzinger, `cond-mat/0501663` (2005).
[5] P.F. Stadler, A. Wagner and D.A. Fell, Adv. Complex Sys. **4**, 207-226 (2001).
[6] M.E.J. Newman, Social Networks **25**, 83-95 (2003).
[7] P. Holme, C.R. Edling and F. Liljeros, Social Networks **26**, 155 (2004).
[8] P. Holme, F. Liljeros, C.R. Edling and B.J. Kim, Phys. Rev. E **68**, 056107 (2003).
[9] M.E.J. Newman, Social Networks **25**, 83 (2003).
[10] M.C. González, P.G. Lind and H.J. Herrmann, "Model of mobile agents for sexual interactions networks", submitted, 2005.
[11] G. Caldarelli, R. Pastor-Satorras and A. Vespignani, Eur. Phys. J. B **38**, 183-186 (2004).
[12] E. Marinari and R. Monasson, J. Stat. Mech., P09004 (2004).
[13] H.D. Rozenfeld, J.E. Kirk, E.M. Bollt and D. ben-Avraham, `cond-mat/0403536` (2004).
[14] G. Bianconi and A. Capocci, Phys. Rev. Lett. **90**, 078701 (2003).
[15] G. Bianconi, G. Caldarelli and A. Capocci, `cond-mat/0408349` (2004).
[16] C.P. Herrero, Phys. Rev. E **71**, 016103 (2005).
[17] S.-J. Yang, Phys. Rev. E **71**, 016107 (2005).
[18] G. Bianconi and M. Marsili, `cond-mat/0502552` (2005).
[19] A. Vázquez, J.G. Oliveira and A.-L. Barabási, Phys. Rev. E **71**, 025103(R) (2005).
[20] M.E.J. Newman and D.J. Watts, Phys. Rev. E **60**, 7332-7342 (1999).
[21] J.S. Andrade Jr., H.J. Herrmann, R.F.S. ANDRADE, L.R. da Silva, Phys. Rev. Lett. **94**, 018702 (2005).
[22] S.N. Dorogovtsev, A.V. Goltsev and J.F.F. Mendes, Phys. Rev. E **65**, 066122 (2002).
[23] J.L. Wylie and A. Jolly, Sex. Transm. Dis. **28**, 14 (2001).
[24] J.J. Potterat, L. Phillips-Plummer, S.Q. Muth, R.B. Rothenberg, D.E. Woodhouse, T.S. Maldonado-Long, H.P. Zimmerman, J.B. Muth, Sex. Transm. Infect. **78**, i159 (2002).
[25] F. Liljeros, C.R. Edling, L.A.N. Amaral and H.E. Stanley, Nature **411**, 907 (2001).
[26] S. Wuchty, E. Ravasz and A.-L. Barabasi, "The Architecture of Biological Networks", in T.S. Deisboeck, J. Yasha Kresh and T. B. Kepler (eds.), *Complex Systems Science in Biomedicine* (Kluwer Academic Publishing, New York, 2003).
[27] H.-J. Kim and J.M. Kim, `cond-mat/0503168`, 2005.
[28] T. Petermann and P. de los Rios, `cond-mat/0401434`, 2004.