Eur. Phys. J. B **38**, 183–186 (2004)
DOI: 10.1140/epjb/e2004-00020-6

THE EUROPEAN PHYSICAL JOURNAL B

Structure of cycles and local ordering in complex networks

G. Caldarelli[1], R. Pastor-Satorras[2,a], and A. Vespignani[3]

[1] INFM UdR Roma 1, Dipartimento di Fisica, Università "La Sapienza", P.le A. Moro 2, 00185 Roma, Italy
[2] Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, 08034 Barcelona, Spain
[3] Laboratoire de Physique Théorique (UMR du CNRS 8627), Bâtiment 210, Université de Paris-Sud, 91405 Orsay Cedex, France

Received 6 November 2003
Published online 17 February 2004 –

Abstract. We study the properties of quantities aimed at the characterization of grid-like ordering in complex networks. These quantities are based on the global and local behavior of cycles of order four, which are the minimal structures able to identify rectangular clustering. The analysis of data from real networks reveals the ubiquitous presence of a statistically high level of grid-like ordering that is non-trivially correlated with the local degree properties. These observations provide new insights on the hierarchical structure of complex networks.

PACS. 89.75.-k Complex systems – 89.75.Fb Structures and organization in complex systems

Empirical evidence shows that the topology of most networks arising in the biological, social, and technological contexts exhibits complex features which cannot be explained by merely extrapolating the local properties of their constituents [1,2]. The most relevant among these features is the small-world property [3] and a high level of heterogeneity, usually reflected in a scale-free behavior of the network's connectivity [4]. While these properties would point to a very large degree of randomness, real networks exhibit a surprising level of structural order. This fact has been first pointed out by noting the common property of many networks to form cliques in which every element is linked to every other element; i.e. the presence of a high clustering coefficient [3]. The identification of hidden ordering and hierarchies in the seemingly haphazard appearance of real networks is therefore a major area of study, aimed at understanding their basic organizing principles. This activity has led to a harvest of results concerning nontrivial correlation properties among the various elements of natural networks, suggesting the presence of interesting modular organizations [5–8].

In this paper we point out that the usual clustering coefficient is in some cases unable to quantify the order underlying a network's structure. In particular, a general ordered network structure is represented by a grid-like frame, such as a regular hypercubic lattice, that can be adequately quantified only by evaluating the frequency of rectangular loops appearing in the network. We introduce a grid coefficient that allows us to uncover the presence of a surprising level of grid ordering in several real networks ranging from technological (the physical Internet) to social (scientific collaboration network) systems. By correlating the presence of grid-like structures with the local connectivity properties we are able to uncover the presence of a hierarchy that appears to be a widely present organizing principle [6,8]. In some cases, the scaling behavior of the grid clustering is very similar to that of the clustering coefficient, suggesting a kind of statistical self-similarity in the modular construction of the network.

A network or graph [9] is a set of vertices and edges joining pairs of vertices, representing individuals and the interactions among them, respectively. Two features play a special role in the characterization of complex networks. The first one refers to the *small-world* concept [3]: i.e. the small average distance in terms of number of edges between any two vertices in the system. The second consists in a very high heterogeneity, usually reflected in a *scale-free* degree distribution $P(k) \sim k^{-\gamma}$ for the probability that any given vertex has degree k; i.e. k edges to other vertices [4]. Both properties appear to be ubiquitous in dynamically growing networks [1,2]. Real networks also show a large degree of local clustering and correlations. A first quantitative measurements of these properties is provided by the clustering coefficient [3]. In particular, the clustering coefficient c_i of the vertex i, with degree k_i, is defined as the ratio between the number of edges e_i in the subgraph identified by its nearest neighbors and its maximum possible value, $k_i(k_i-1)/2$, corresponding to a complete sub-graph, i.e. $c_i = 2e_i/k_i(k_i-1)$. The average clustering coefficient $\langle c \rangle$ is defined as the average value of c_i over all the vertices in the graph, $\langle c \rangle = \sum_i c_i/N$, where N is the size of the network. This magnitude quantifies the relative

[a] e-mail: romualdo.pastor@upc.es

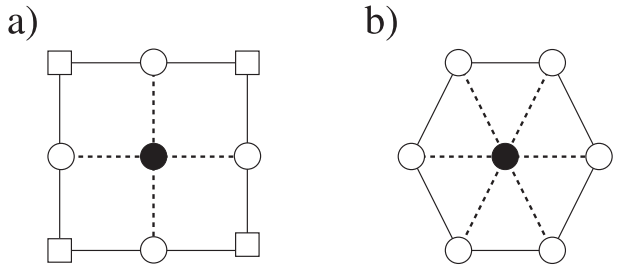

Fig. 1. (a) Regular square lattice. Nearest neighbors of a vertex (empty circles) are not neighbors of each other. Therefore the clustering coefficient $c_i \equiv 0$ for every vertex i. (b) Triangular lattice. Here some of the neighbors are connected to each other. In particular 2 out of every 5 possible edges are drawn; hence $c_i = 2/5$ for all the vertices.

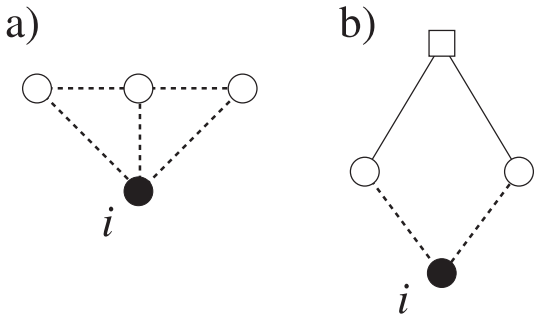

Fig. 2. (a) Example of a primary quadrilateral, in which the three external vertices are nearest neighbors of the vertex i. (b) Example of a secondary quadrilateral in which one of the external vertices (empty square) is a second neighbor of the vertex i .

abundance with which two vertices connected to the same vertex are also connected to each other. By comparison, random graphs [10] are not clustered, having $\langle c\rangle = \langle k\rangle/N$, where $\langle k\rangle$ is the average degree, while triangular lattices tend to be highly clustered with their neighbors.

Further information can be extracted if one computes the average clustering coefficient $c(k)$ as a function of the vertex degree k [6]. In the physics terminology, the study of the clustering coefficient $c(k)$ is strictly related to the analysis of three-point correlation functions [11]. The absolute average value – as well as the scaling with k – of this quantity are fundamental to discriminate the level of randomness and the organizing principles related to the basic hierarchies present in the networks. For instance, a large class of scale-free networks shows a clustering coefficient decaying as a power-law as a function of the vertex's degree [8]. This implies that low degree vertices tend to form connected cliques with other vertices, while large connected vertices (hubs) tend to act as bridges between unconnected cliques, thus showing a small clustering coefficient. This fact highlights the existence of some modular building, identified by the cliques of small degree vertices [8].

With the aim of unveiling the hidden ordering in complex networks, the use of the two- and three-point correlations is however not always sufficient. As a very simple example we can consider a rectangular lattice or grid, Figure 1a. In this case it is easy to recognize that the clustering coefficient is not able to distinguish any architecture in a grid-like structure, since its value is always null. However, it is a good measure of order for other regular structures, such as a triangular lattice, Figure 1b. Since grid-like structures are quite frequently observed patterns in natural systems, we introduce as a further quantitative characterization of networks' regularity some metrics that naturally account for rectangular symmetries [12–15].

We start by considering the closed paths in a network in which all edges and vertices are distinct. These closed paths are known as cycles [9]. Cycles of length 3 (i.e. composed of three vertices) are called *triangles*. The ratio between the number of triangles that include the vertex i, e_i, and its maximum possible number, $k_i(k_i-1)/2$, defines the triangle coefficient of the vertex i, which is by definition equal to its clustering coefficient c_i. Cycles of length 4 are called *quadrilaterals*. In the spirit of the clustering coefficient, we want to improve the measurement of the network structure by using the *grid coefficient*, $c_{4,i}$, that is defined as the fraction of all the quadrilaterals passing by the vertex i, Q_i, divided by the maximum possible number of quadrilaterals sharing the vertex i, Z_i. Analogously, one could consider cycles of of length n, and define the corresponding coefficient $c_{n,i}$ as the fraction of of all the cycles of length n that pass through the vertex i, divided by the maximum number of those cycles that could pass by i. The computational effort to calculate $c_{n,i}$ grows quite fast with n. Therefore in the present work we will focus in the simplest nontrivial case $n = 4$.

The grid coefficient defined for cycles of length 4 can be further decomposed by noting that each quadrilateral passing by i is composed of the vertex i itself plus three external vertices. Quadrilaterals can be therefore classified according to the nature of the external vertices, see Figure 2. If all the external vertices are nearest neighbors of i, they form a *primary quadrilateral*; on the other hand, if one of the external vertices is a second neighbor of i, the cycle they form is a *secondary quadrilateral*. If the vertex i has degree k_i and it is connected to $k_{i,2\mathrm{nd}}$ second neighbors, it is easy to check that the maximum number of primary quadrilaterals is $Z_i^p = 3\times\binom{k_i}{3} = k_i(k_i-1)(k_i-2)/2$, while the maximum number of secondary quadrilaterals is $Z_i^s = k_{i,2\mathrm{nd}}k_i(k_i-1)/2$. In this way, in order to study the grid properties of a network, we can define three magnitudes: the primary grid coefficient, $c_{4,i}^p = Q_i^p/Z_i^p$, the secondary grid coefficient $c_{4,i}^s = Q_i^s/Z_i^s$, and the total grid coefficient $c_{4,i} = (Q_i^p + Q_i^s)/(Z_i^p + Z_i^s)$, where Q_i^p and Q_i^s are the actual number of primary and secondary quadrilaterals passing by the node i, respectively. The respective average grid coefficients are defined by averaging these quantities over all vertices in the network and define the global relative abundance of quadrilaterals in the network.

As an example of this definition, let us consider the rectangular lattice represented in Figure 1a, in which each vertex i has 4 nearest neighbors and 8 second neighbors. There are no primary quadrilaterals passing by any node i, while the number of secondary quadrilaterals is

$Q^s = 4$. From here we obtain $\langle c_4^p \rangle = 0$, $\langle c_4^s \rangle = 1/9$, and $\langle c_4 \rangle = 1/15$. On the other hand, in the triangular lattice, Figure 1b, in which each vertex has 6 nearest neighbors and 12 second neighbors, we find 6 primary quadrilaterals and 6 secondary quadrilaterals, which yield $\langle c_4^p \rangle = 1/10$, $\langle c_4^s \rangle = 1/30$, and $\langle c_4 \rangle = 1/20$. Thus, regular grids exhibit a finite grid coefficient, in opposition to the clustering coefficient, which is zero for any hypercubic lattice.

A very different case is represented by a random network with fixed degree distribution, an example of which is given by the configuration model [15,16]. For a random network, the probability that a randomly chosen edge points to a vertex of degree k is $q(k) = kP(k)/\langle k \rangle$. On the other hand, the probability that two vertices of degrees k_i and k_j are connected is $\pi(k_i, k_j) = k_i k_j / \langle k \rangle N$. For any vertex i, we need at least three nearest neighbors to construct a primary quadrilateral. Given this configuration, the probability to close the cycle in any of the three possible quadrilaterals is given by the probability to draw two edges between two of the three nearest neighbors. Therefore, we have a primary grid coefficient $\langle c_4^p \rangle_{\rm RG} = \sum_{k_i,k_j,k_l} q(k_i)\pi(k_i-1,k_j-1)q(k_j)\pi(k_j-2,k_l-1)q(k_l) = (\langle k^2 \rangle - \langle k \rangle)^2(\langle k^3 \rangle - 3\langle k^2 \rangle + 2\langle k \rangle)/(\langle k \rangle^5 N^2)$. This implies that a random graph with finite $\langle k^2 \rangle$ and $\langle k^3 \rangle$, has an average primary grid coefficient $\langle c_4^p \rangle_{\rm RG} \sim N^{-2}$. The calculation for the secondary grid coefficient is slightly more involved. In this case, for any vertex i, we need at least two nearest neighbors and a second neighbor. This last vertex, being a second neighbor, is connected to at least one nearest neighbor, but not necessarily to any of the two nearest neighbors that will compose the quadrilateral. If the second neighbor is not a priori connected to the two nearest neighbors, then the probability of finding a quadrilateral is of order N^{-2}. On the other hand, if it is a priori connected to one of the selected nearest neighbors, the probability of closing a quadrilateral is given by $\sum_{k_j k_l} q(k_j)\pi(k_j - 1, k_l - 1)q(k_l) = (\langle k^2 \rangle - \langle k \rangle)^2/(\langle k \rangle^3 N) \equiv \langle c \rangle_{\rm RG}$, which coincides with the general expression for the clustering coefficient [15]. This last instance (that the second neighbors is a priori connected to one of the nearest neighbors considered) happens with probability $1/k_i$, where k_i is the degree of the vertex i. Therefore, at leading order in N^{-1}, we have that the average secondary grid coefficient in a random graph is given by $\langle c_4^s \rangle_{\rm RG} = \langle c \rangle_{\rm RG} \sum_{k \geq 2} P(k)/k$. For a random graph with a bounded degree distribution with finite moments, we have that the grid coefficient scales as $\langle c_4 \rangle_{\rm RG} \sim N^{-1}$ with the number of vertices N. For a scale-free random graph, on the other hand, the degree moments can be large, and yield therefore non-vanishing grid coefficients even for large N. It is also worth noticing that in the case of $\gamma < 7/3$ the configuration model gives unphysical results due to the presence of double edges and loops [17].

In order to characterize the level of grid-like ordering in real networks, we have measured the grid coefficients in four different systems, characterized by a scale-free degree distribution:

Internet: Internet map at the Autonomous System (AS) level, as of 22nd November 1999 [5,6,18]. These maps are collected and made publicly available by the National Laboratory for Applied Network Research (NLANR)[1]. Each AS refers to one single administrative domain of the Internet. Different ASs are in most cases connected through a Border Gateway Protocol (BGP) that identifies any AS through a 16-bit number. The map considered is composed of 6243 ASs acting as vertices and by 12113 BGP peer connections, acting as edges, yielding an average degree $\langle k \rangle = 3.88$.

World-Wide-Web: Map of the World-Wide-Web collected at the domain of Notre Dame University[2] [19–21]. This network is actually directed, but we have considered it as non-directed. The map is composed of 325729 web pages, represented by vertices, and 1090108 hyperlinks pointing from one page to another, represented by edges, which corresponds to an average degree $\langle k \rangle = 6.69$.

Yeast protein map: Protein interaction map of the yeast *Saccharomyces Cerevisiae*[3] [22,23]. This network is composed of 2874 proteins, that constitute the vertices, and 7753 protein-protein interactions, identified by two amino-acid chains binding to each other, that constitute the edges, for an average degree $\langle k \rangle = 5.40$.

Scientific collaborations: Network of scientific collaborations collected from the condensed matter preprint database at Los Alamos[4] [24,25]. The graph is composed of 16264 different authors, that are connected by one edge if they have coauthored a joint paper. The total amount of collaborations (edges) is then 47594, yielding an average degree $\langle k \rangle = 5.85$.

In Table 1 we report the different average grid coefficients for all the networks analyzed, compared with those corresponding to a random graph with the same size and degree distribution. It is interesting to note that, with the exception of the Internet, in which the random graph configuration model gives unphysical results [17], the average grid coefficients in most networks are one to four orders

Table 1. Average degree, primary, secondary, and total grid coefficients for the different networks considered, compared with the theoretical values for a random networks with the same size, average degree and degree distribution (see text).

	Internet	WWW	yeast	cond-mat
$\langle k \rangle$	3.88	6.69	5.40	5.85
$\langle c_4^p \rangle$	0.043	0.14	0.021	0.40
$\langle c_4^p \rangle_{\rm RG}$	5.95	0.021	0.005	5×10^{-6}
$\langle c_4^s \rangle$	0.028	0.088	0.008	0.036
$\langle c_4^s \rangle_{\rm RG}$	0.24	0.004	0.007	3×10^{-4}
$\langle c_4 \rangle$	0.028	0.090	0.010	0.12

[1] The NLANR is sponsored by the National Science Foundation (see `http://moat.nlanr.net/`).

[2] Data publicly available at `http://www.nd.edu/~networks`.

[3] Data available at the DIPTM database `http://dip.doe-mbi.ucla.edu`

[4] Database located at `http://xxx.lanl.gov/archive/cond-mat`

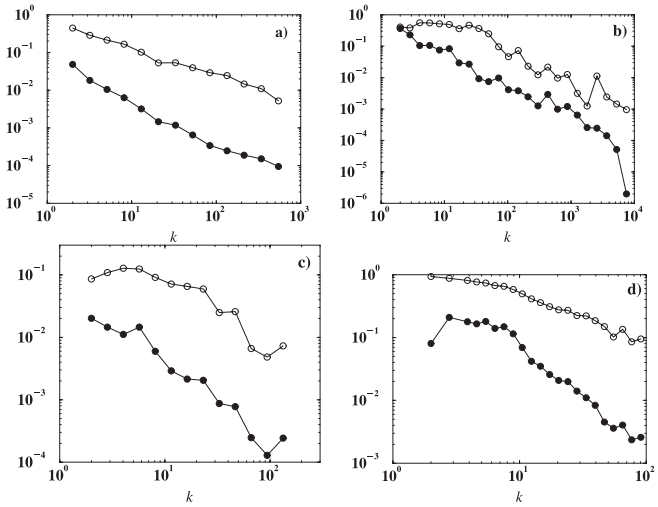

Fig. 3. Clustering coefficient $c(k)$ (hollow symbols) and grid coefficient $c_4(k)$ (filled symbols) as a function of the degree, for the networks considered. (a) Internet at the AS level. (b) Map of the World-Wide-Web domain collected at www.nd.edu. (c) Network of protein interactions in the yeast *Saccharomyces Cerevisiae*. (d) Scientific collaborations from the cond-mat preprint database.

of magnitude larger than the corresponding coefficients of a random graph. While the small-world property and the scale-free degree distribution common to all these networks are generally associated to disorder and large fluctuations, the presence of large grid coefficient makes those graphs reminiscent of a grid-like ordering.

More information can be gathered by studying the grid coefficient as a function of the vertex's degree k (i.e. by considering the average value $c_4(k)$ of the total grid coefficient for all the vertices with the same degree k). As similarly noticed for the clustering coefficient [6,8], the grid coefficient is well approximated in most cases by a power-law decay for increasing k. This feature indicates a correlation between the vertices' degree and the local network structure. In particular, low degree vertices are arranged in fairly ordered patterns whose building blocks are triangular and rectangular structures. Vertices with large degree act as the network backbone by connecting the highly clustered regions. Since we are facing power-law behavior for the clustering and grid coefficients, we have that no characteristic length scales are present in the system and thus there is a hierarchy of modular structures incorporating loops of all lengths, appearing at different length scales. Even though statistical fluctuations are comparable, in some cases the grid coefficient appears to be less susceptible to noise than other metrics. Finally, we note the apparent presence of two classes of networks: the first with a scaling of the $c_4(k)$ very similar to $c(k)$ (corresponding to the Internet and the WWW), and a second one with $c_4(k)$ different from $c(k)$ (the protein and scientific collaboration maps). This observation can be interpreted as follows: When the power-law behavior is alike, we can talk of *self-similar networks* in which both rectangular and triangular patterns are equally implemented in the modular construction of the network. In the second situation, one of the two patterns is abandoned earlier in the hierarchical construction of the graph, breaking the self-similarity of the hierarchy.

The authors thank M.E.J. Newman for making available his data sets on scientific collaborations. This work has been partly supported by the EC-Fet Open project COSIN IST-2001-33555. R.P.-S. acknowledges the support from the Ministerio de Ciencia y Tecnología (Spain) and from the Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya (Spain).

References

1. R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
2. S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002)
3. D.J. Watts, S.H. Strogatz, Nature **393**, 440 (1998)
4. A.-L. Barabási, R. Albert, Science **286**, 509 (1999)
5. R. Pastor-Satorras, A. Vázquez, A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001)
6. A. Vázquez, R. Pastor-Satorras, A. Vespignani, Phys. Rev. E **65**, 066130 (2002)
7. M.E.J. Newman, Phys. Rev. Lett. **89**, 208701 (2002)
8. E. Ravasz, A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003)
9. B. Bollobás, *Modern Graph Theory* (Springer-Verlag, New York, 1998)
10. P. Erdös, P. Rényi, Publicationes Mathematicae **6**, 290 (1959)
11. A. Vázquez, R. Pastor-Satorras, A. Vespignani, Internet topology at the router and autonomous system level, 2002, e-print `cond-mat/0206084`
12. A. Vázquez, A. Flammini, A. Maritan, A. Vespignani, ComPlexUs **1**, 38 (2003)
13. P. Holme, C.R. Edling, F. Liljeros, *Structure and time-evolution of the Internet community pussokram.com, 2002,* e-print `cond-mat/0210514`
14. G. Bianconi, A. Capocci, Phys. Rev. Lett. **90**, 078701 (2003)
15. M.E.J. Newman, in *Handbook of Graphs and Networks: From the Genome to the Internet*, edited by S. Bornholdt, H.G. Schuster (Wiley-VCH, Berlin, 2003), pp. 35–68
16. M. Molloy, B. Reed, Random Struct. Algorithms **6**, 161 (1995)
17. M.E.J. Newman, J. Park, Phys. Rev. E **68**, 036122 (2003)
18. M. Faloutsos, P. Faloutsos, C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999)
19. R. Albert, H. Jeong, A.-L. Barabási, Nature **401**, 130 (1999)
20. A.-L. Barabási, R. Albert, H. Jeong, Physica A **281**, 69 (2000)
21. B.A. Huberman, L.A. Adamic, Nature **401**, 131 (1999)
22. A. Wagner, Mol. Biol. Evol. **18**, 1283 (2001)
23. H. Jeong, S.S. Mason, A.L. Barabási, Z.N. Oltvai, Nature **411**, 41 (2001)
24. M.E.J. Newman, Phys. Rev. E **64**, 016131 (2001)
25. A.-L. Barabási et al., Physica A **311**, 590 (2002)