Physica A 392 (2013) 2576–2587

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# A network approach based on cliques

I.S. Fadigas [a], H.B.B. Pereira [b,c,*]

[a] Departamento de Ciências Exatas, Universidade Estadual de Feira de Santana, Campus Universitário, Módulo 5, 44031-460, Feira de Santana, BA, Brazil
[b] Programa de Modelagem Computacional, SENAI Cimatec, Av. Orlando Gomes 1845, 41.650-010, Salvador, BA, Brazil
[c] Departamento de Educação, Universidade do Estado da Bahia, Rua Silveira Martins, 2555, 41.150-000, Salvador, BA, Brazil

ARTICLE INFO

*Article history:*
Received 4 May 2012
Received in revised form 17 November 2012
Available online 1 February 2013

*Keywords:*
Networks of cliques
Complex networks
Small-world phenomenon
Social network
Semantic networks

ABSTRACT

The characterization of complex networks is a procedure that is currently found in several research studies. Nevertheless, few studies present a discussion on networks in which the basic element is a clique. In this paper, we propose an approach based on a network of cliques. This approach consists not only of a set of new indices to capture the properties of a network of cliques but also of a method to characterize complex networks of cliques (i.e., some of the parameters are proposed to characterize the small-world phenomenon in networks of cliques). The results obtained are consistent with results from classical methods used to characterize complex networks.



## 1. Introduction

*Network of Cliques* is a type of network that has applications to real situations (e.g. film actors' networks, co-authorship network, and semantic networks based on written and oral discourses). The basic element in a network of cliques is not a vertex but instead is a set of $n$ vertices that are mutually connected, i.e., a clique. These networks have a specific characteristic: their formation is based on juxtaposition and/or an overlapping of cliques. On the one hand, a *juxtaposition process* means to connect two cliques with only one common vertex. On the other hand, when the connection between cliques occurs with two or more vertices, we have an *overlapping process*.

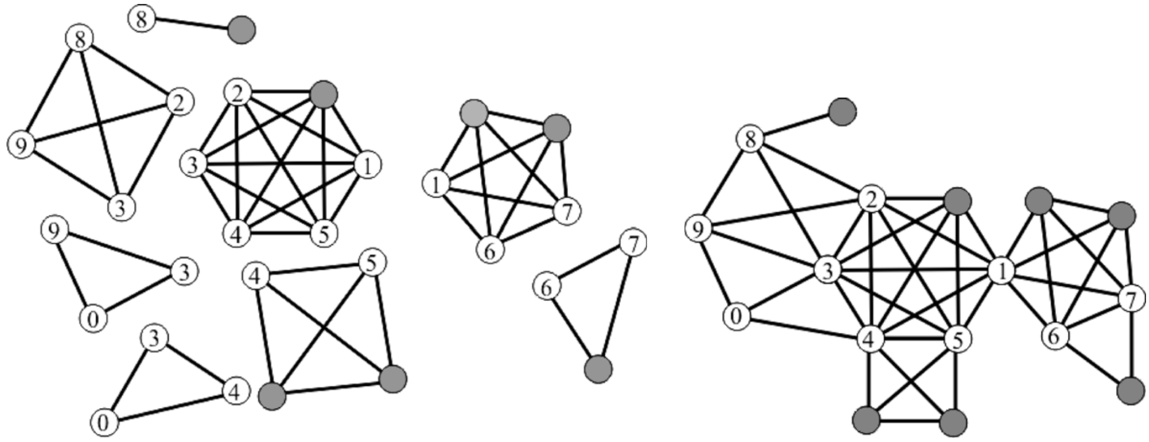
'Networks of cliques' are represented as a graph $G = (V, \mathcal{E})$ that is a mathematical structure and consists of two sets: $V$ (finite and not empty) and $\mathcal{E}$ (a binary relation on $V$). The elements of $V$ are called vertices, and the elements of $\mathcal{E}$ are called edges [1]. In our networks of cliques, each edge has two vertices associated with it. If $V_G$ is the set of vertices of the graph $G$, then a maximal subset of $V_G$ is called a *clique* if each pair of vertices is connected by one edge.

A highly treated example of 'network of cliques' is the film actors' network [2–5]. In this network, each cast of actors builds a clique, and each clique of actors is connected with another if they have an actor, or actors, in common. Another example is that of semantic networks based on titles of scientific papers, in which the cliques are built by words in titles, and two titles are connected if they have words in common [6,7].

Fig. 1(a) shows an example of an initial state of isolated cliques that consists of $t$ cliques before the juxtaposition and/or overlapping processes. Fig. 1(b) depicts an example of a 'network of cliques' after the juxtaposition and/or overlapping processes. 'Networks of cliques' have different properties in comparison with other types of networks because of their building process.

* Corresponding author at: Programa de Modelagem Computacional, SENAI Cimatec, Av. Orlando Gomes 1845, 41.650-010, Salvador, BA, Brazil. Tel.: +55 7187411321.
*E-mail addresses:* isfadigas@gmail.com (I.S. Fadigas), hbbpereira@gmail.com (H.B.B. Pereira).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2013.01.055




(a) Initial state of isolated cliques.

(b) A 'network of cliques' connected by juxtaposition and/or by overlapping.

**Fig. 1.** An initial state of isolated cliques and a possible 'network of cliques'.

This paper is organized as follows. In Section 2, we introduce four theoretical structures of 'networks of cliques' that are minimally connected, on which later discussions will be based. In Section 3, we discuss some properties based on measures of group cohesion from a theoretical perspective. An empirical analysis is performed in Section 4, in which we use semantic networks that are based on titles of scientific papers with a large number of vertices, so that we can study properties that emerge from such networks. Finally, our concluding remarks are presented in Section 5.

## 2. Theoretical structures of 'networks of cliques'

We define a *'network of cliques' that are minimally connected* as one in which each clique is connected with another by a juxtaposition process. This definition includes some topologies or settings for this type of network (Fig. 2).

A 'network of cliques' is said to be minimally connected in a line layout if no clique is connected to more than two others (Fig. 2(a)). Specifically, when the line layout is closed (i.e., if each clique is minimally connected with exactly two others), the 'network of cliques' is said to be minimally connected in a circle layout (Fig. 2(b)). A 'network of cliques' is said to be minimally connected in a star layout if all of the cliques are connected by only one common vertex (Fig. 2(c)). These configurations are analogous to the classical types of graphs: line, circle and star, respectively (Fig. 3(a)–(c)).

Another configuration, called 'network of cliques', minimally connected and arranged in a layer layout (Fig. 2(d)), is defined as follows: in the central position is a clique and other cliques are connected with it by its 'free' vertices (i.e., each clique is connected with the central clique by each vertex of the central clique that is not connected yet). The process of "free" vertex connections continues until the last clique is connected. In a specific case, when all of the cliques are dyads, then there is no distinction between the line and layer layouts.

In the next section, we use the 'networks of cliques', minimally connected, especially the line and star layout (Fig. 2(a) and (c), respectively), as theoretical structures with which some real networks will be compared. In addition, we will analyze the results of some measures of cohesion from the number of vertices and edges, the number of cliques and the sizes of the cliques.

## 3. Indices of cohesion

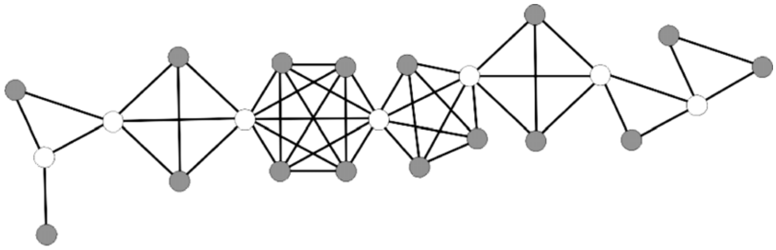
Seven concepts and measures of group cohesion are accounted for to establish a theoretical foundation: the density, average degree, fragmentation, cut points, geodesic distance, diameter, and clustering coefficient.

### 3.1. Density

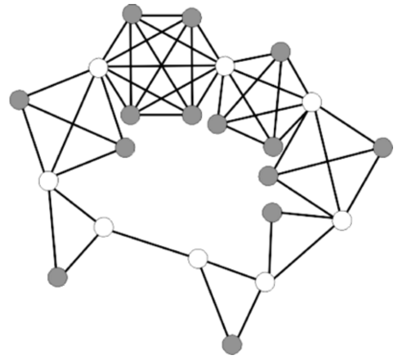
Among the indices of cohesion of a network, one of the most important indices is the density ($\Delta$). The *density* of an undirected network is the total of the existing edges ($|\mathcal{E}|$) divided by the maximum possible number of edges ($n(n-1)/2$). The classical mathematical expression for calculating the density for undirected networks is given in Eq. (1).

$$\Delta = \frac{2|\mathcal{E}|}{n(n-1)}. \tag{1}$$

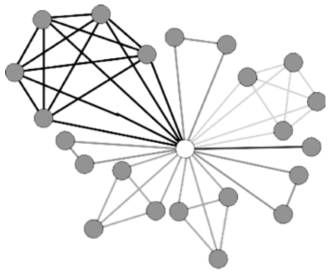
For networks in general, the density varies between 0 and 1. The value 0 occurs when the network is completely unconnected, i.e., all of the vertices are isolated. Because $n-1$ is the minimum number of edges for a network to be connected,



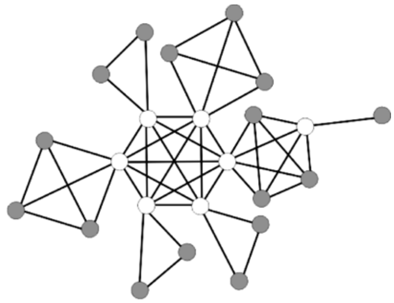
(a) 'Network of cliques', minimally connected and arranged in a line layout.



(b) 'Network of cliques', minimally connected and arranged in a circle layout.



(c) 'Network of cliques', minimally connected and arranged in a star layout.



(d) 'Network of cliques', minimally connected and arranged in a layer layout.

**Fig. 2.** Examples of theoretical structures of 'networks of cliques' that are minimally connected by juxtaposition process. The white vertices are cut vertices (or cut points) in the line, star and layer layouts, or they are the connection vertices (or connection points) in a circle layout.

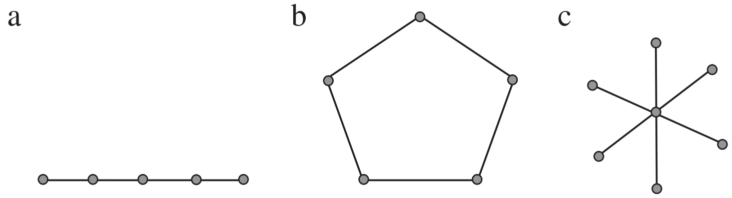



**Fig. 3.** Types of graphs or networks: (a) line, (b) circle and (c) star.

the density of such a network is $2/n$. Given a set of vertices, connecting the vertices with the minimum number of edges corresponds to building a 'network of cliques' formed by dyads that are minimally connected.

Consider a 'network of cliques' that is nontrivial, i.e., one in which the size of a clique is greater than or equal to 3. When we have an initial state of isolated cliques (Fig. 1(a)), the number of edges is given by the sum of the quantity of edges in each clique. Let $n_q$ denote the number of cliques, $q_i$ the size (i.e., the quantity of vertices) of the $i$th clique, and $n_0$ the total number of vertices of an initial state of isolated cliques; then, the density of the initial state of isolated cliques can be expressed by the following:

$$\Delta_{q0} = \frac{2|\mathcal{E}|}{n_0(n_0-1)} = \frac{\sum_{i=1}^{n_q} q_i(q_i-1)}{n_0(n_0-1)}. \tag{2}$$

For an initial state of isolated cliques, $\Delta_{q0} \neq 0$, once $|\mathcal{E}| \neq 0$. Then, $\Delta_{q0}$ must be used to normalize the classical density (Eq. (1)) of the 'network of cliques' (i.e., the network that results from the juxtaposition and/or overlapping process). The normalized density of an initial state of isolated cliques is given by Eq. (3).

$$\Delta_{norm} = \frac{\Delta - \Delta_{q0}}{1 - \Delta_{q0}}. \tag{3}$$

Interpreting the above definition (Eq. (3)), the normalized density $\Delta_{norm} = 0$ corresponds to the case where there is no juxtaposition or overlapping process (i.e. there is no common vertices to cliques in a network) and $\Delta = \Delta_{q0}$. The normalized density $\Delta_{norm} = 1$ corresponds to the case where the overlapping process is complete and the resulting network is a single clique and $\Delta = 1$. Thus, the normalized density zero means that the cliques are not affected by the juxtaposition or overlapping process.

Let $\upsilon$ express variations. An expression to quantify the density variation of the 'network of cliques', compared with its initial state of isolated cliques (Fig. 1(a)), is given in Eq. (4):

$$\upsilon(\Delta) = \frac{\Delta - \Delta_{q0}}{\Delta_{q0}}. \tag{4}$$

The advantage of using this expression is that we can measure the density variation related to $\Delta_{q0}$, i.e., the greater the variation is, the more cohesive is the 'network of cliques' with respect to the initial state (i.e., the initial state of isolated cliques). Thus, it could be that, for the same values of $\Delta$ in different networks, different values of $\upsilon(\Delta)$ are obtained. Eq. (4) can be written in terms of edges and vertices, replacing the densities by their expressions, which results in Eq. (5).

$$\upsilon(\Delta) = \frac{|\mathcal{E}|}{|\mathcal{E}_0|} \cdot \frac{n_0}{n} \cdot \frac{n_0 - 1}{n - 1} - 1 \tag{5}$$

where $n_0$ represents the number of vertices of an initial state of isolated cliques and $n$ is the number of vertices of a 'network of cliques' after running a juxtaposition and/or overlapping process. Similarly, the number of edges follows the same assumptions. For a real 'network of cliques', $n \gg 1$ and $n_0 \gg 1$; therefore, we can use $n - 1 \simeq n$ and $n_0 - 1 \simeq n_0$. With these approximations, Eq. (5) can be written as follows:

$$\upsilon(\Delta) \simeq \frac{|\mathcal{E}|}{|\mathcal{E}_0|} \cdot \frac{n_0^2}{n^2} - 1 = \frac{[n_0/n]^2}{[|\mathcal{E}_0|/|\mathcal{E}|]} - 1. \tag{6}$$

Then, the density variation is directly proportional to the square of the variation in the number of vertices and inversely proportional to the variation in the number of edges.

### 3.2. Average degree

The *average degree* of an undirected network is the average of $k_i$, which is denoted by $\langle k \rangle = \frac{1}{n} \sum_i^n k_i$, where $n = |V|$. Another way to denote the *average degree* of an undirected network is $\langle k \rangle = 2|\mathcal{E}|/n$. Therefore, the average degree is associated with the density by Eq. (7).

$$\Delta = \frac{\langle k \rangle}{n - 1} = \frac{\langle k \rangle}{k_{\max}}. \tag{7}$$

Therefore, the average degree for the structure of the 'network of cliques' presented earlier (Fig. 2) is expressed by Eq. (8).

$$\langle k_{q0} \rangle = \frac{\sum_{i=1}^{n_q} q_i(q_i - 1)}{n_0}. \tag{8}$$

From the definition above, we can establish a parameter for the 'network of cliques' that captures the variation rate of the average degree, i.e., a parameter that compares the average degree of the initial state of isolated cliques with the average degree of the real 'network of cliques'. Eq. (9) shows this variation.

$$\upsilon(\langle k \rangle) = \frac{\langle k \rangle - \langle k_{q0} \rangle}{\langle k_{q0} \rangle}. \tag{9}$$

### 3.3. Fragmentation

The *fragmentation* of a network accounts for the relationship between the number of components and the number of vertices. A definition of fragmentation that considers the size of the components [8] is given by Eq. (10).

$$F = 1 - \frac{\sum_{j=1}^{m} n_j(n_j - 1)}{n(n - 1)} \tag{10}$$

where $n_j$ is the number of vertices of the component $j$ and $m$ is the number of components of the fragmented network. For an unfragmented network, $F = 0$, and for a completely fragmented network, $F = 1$. If $0 < F \leq 1$, then the network is unconnected.

For the initial state of isolated cliques, the maximum should occur when the cliques are all unconnected, as in Fig. 1(a). We propose that the fragmentation index for the 'network of cliques' is given by Eq. (11).

$$F_{clique} = \frac{Comp - 1}{n_q - 1}. \tag{11}$$

In this equation, when the quantity of components (*Comp*) is exactly the number of cliques in an initial state of isolated cliques, $F = 1$ (i.e., the 'network of cliques' is completely fragmented). On the other hand, when the 'network of cliques' is connected ($Comp = 1$), $F = 0$. This index measures a fragmentation based on the reduction of the components (i.e., the defragmentation process during the formation of the 'network of cliques'). In this approach, the size of the components is not accounted for, i.e., what matters is how the cliques are connected regardless of their initial size.

### 3.4. Cut point

From a connected network, we can remove some vertices that produce an unconnected network (i.e., a network with two or more components). These special vertices are called *cut points* ($P_c$). The quantity of cut points depends on the topology of the network. For example, for the same quantity $n$ of vertices, the star network (Fig. 3(c)) has only one cut point, while for the line network (Fig. 3(a)), it has $n - 2$ cut points, and for the circle network (Fig. 3(b)), it has no cut points.

For a 'networks of cliques' to be minimally connected, the maximum number of cut points occurs in the 'networks of cliques' that are arranged in the line or the layer layouts (Fig. 2(a) and (d)), because in these layouts each clique is connected with another by only one vertex, which makes up a total of $n_q - 1$ cut vertices; the other vertices of the cliques are not cut vertices. On the other hand, 'networks of cliques' arranged in a star layout (Fig. 2(c)) has only one cut point, and 'networks of cliques' arranged in a circle layout (Fig. 2(b)) has no cut points. We can then normalize the number of cut points in a real 'network of cliques' using the maximum amount allowed by Eq. (12).

$$Pc_{norm} = \frac{P_c}{n_q - 1}. \tag{12}$$

This index measures the proportion of the maximum possible number of cut points. Low values of $Pc_{norm}$ could be associated, in a non-exclusive way, with both the juxtaposition and the overlapping processes in 'networks of cliques' arranged in circle or star layouts. High values indicate juxtaposition associated with 'networks of cliques' arranged in layer or linear layouts.

### 3.5. Diameter

Although it is not possible to establish a simple expression to calculate the average shortest path or the geodesic distance $l$ to a 'networks of cliques' that are minimally connected in terms of the number of cliques $n_q$, we present methods for calculating the *diameter* of the four 'networks of cliques' structures (Fig. 2):

- For a 'networks of cliques' arranged in a line layout (Fig. 2(a)), the diameter $D = n_q$.
- For a 'networks of cliques' arranged in a circle layout (Fig. 2(b)), the diameter $D = \lceil (n_q + 1)/2 \rceil$, where the notation $\lceil x \rceil$ denotes the smallest integer not less than $x$.
- For a 'networks of cliques' arranged in a star layout (Fig. 2(c)), the diameter $D = 2$.
- Finally, the diameter for a 'networks of cliques' arranged in a layer layout (Fig. 2(d)) depends on the topology and the distribution of the sizes of the cliques. Let $Q_0$ denote the central clique of a 'networks of cliques' arranged in a layer layout. We can call $Q_1$ the first layer of cliques, i.e., the cliques that are connected to a central clique ($Q_0$) by its vertices. Subsequently, $Q_i$ can be understood as an $i$th layer, and the process continues until the last layer $Q_{N_c}$. The index $N_c$ is the total number of layers. In addition, $N_c$ is not the same for a given distribution of sizes of cliques; $N_c$ also depends on how the cliques are connected. Thus, for a 'networks of cliques' arranged in a layer layout, the diameter is given by $D = (2N_c + 1)$, if the cliques of the last layer are connected to different cliques of the penultimate layer, and the diameter $D = 2N_c$ if the cliques of the last layer are connected to only one clique of the penultimate layer.

In summary, the structure of 'networks of cliques' arranged in a line layout maximizes the diameter, while the structure 'networks of cliques' arranged in a star layout minimizes the diameter.

We can then establish an appropriate normalization for the diameter of the 'networks of cliques' with the same number of cliques, to reflect a topology closer to the theoretical structures. When you have $n_q > 3$, a value much lower than expected for real 'networks of cliques', the smaller diameter is, theoretically for 'networks of cliques', arranged in a star layout, followed by the diameter of 'networks of cliques' arranged in circle and linear layouts, in that order. The diameter of the 'networks of cliques' arranged in a layer layout is between the extremes. We define a "*reference diameter*" of 'networks of cliques' that are minimally connected as follows:

$$D_{ref} = \frac{D - 2}{n_q - 2}. \tag{13}$$

The $D_{ref}$ index ranges between 0 and 1 because $n_q > 2$, and the 'networks of cliques' is not a single clique, in which case the diameter is 1. Note that the reference diameter $D_{ref}$ is a cohesion index of cliques, while the geodesic distance $l$ measures the cohesion of the vertices. Similar to the values of $n_q \gg D$ for the 'networks of cliques', the resulting values of Eq. (13) are usually close to zero, which complicates the interpretation. A change of scale is convenient and can be performed with a logarithmic normalization proposed by Eq. (14).

$$D^*_{ref} = \frac{\ln(D/2)}{\ln(n_q/2)}. \tag{14}$$

**Table 1**
Classification of types of theoretical structures of 'networks of cliques' that are minimally connected according to the reference diameter $D^*_{ref}$.

| Reference diameter | Theoretical structures of 'networks of cliques' |
|---|---|
| 0–0.25 | Star layout |
| 0.26–0.75 | Circle or layer layouts |
| 0.76–1.00 | Line layout |

Arbitrarily but consistently, we can establish a classification scale that associates the ranges of reference diameter values with the theoretical structures of 'networks of cliques' that are minimally connected, as shown in Table 1.

### 3.6. Clustering coefficient

The *clustering coefficient* measures the transitivity between the vertices of a network, i.e., the probability that two vertices that are connected to a third vertex are also connected. Eq. (15) [2] is widely used to calculate the clustering coefficient of networks and is characterized by a focus on local indices (i.e., the clustering coefficient of each vertex based on its neighbors, which is given by $C_i = \frac{2\mathcal{E}_i}{k_i(k_i-1)}$).

$$C_{ws} = \frac{1}{n}\sum_{i=1}^{n} C_i. \tag{15}$$

Another approach, which accounts for the overall transitivity, uses the equivalent Eqs. (16) or (17) ([9,5], respectively).

$$C = \frac{3 \times number\ of\ triangles\ in\ the\ network}{number\ of\ connected\ triples\ of\ vertices} \tag{16}$$

$$C = \frac{6 \times number\ of\ triangles\ in\ the\ network}{number\ of\ path\ of\ length\ two}. \tag{17}$$

The intention here is to find a way to express the transitivity of the cliques, from the values of the transitivity of their vertices that are calculated for the theoretical structures of 'networks of cliques' that are minimally connected. The theoretical structures of 'networks of cliques' minimally connected are those that optimize the transitivity of the cliques: (i) the 'networks of cliques' arranged in a star layout maximize the transitivity (in this type of arrangement, all of the cliques are connected to all of the others), and (ii) the 'networks of cliques' arranged in a line layout minimize the transitivity (in this type of arrangement, there is no transitivity of the cliques).

Let us begin by examining the clustering coefficient $C_{ws}$. For the structure of 'networks of cliques' arranged in a star layout, all of the vertices have $C_i = 1$, except for the common vertex to all of the cliques and those vertices of degree ($S_{k_1}$). Therefore, the total contribution of the vertices that have $C_i = 1$ to $\sum C_i$ of Eq. (15) is given by $n - 1 - S_{k_1}$. For the common vertex, the clustering coefficient is less than 1. Thus, $C_{ws} \simeq 1$ for large values of $n$ because in general $S_{k_1} \ll n$ for connected networks. For the 'networks of cliques' that are arranged in a line layout, the fraction of vertices with $C_i = 1$ is given by $n - (n_q - 1) - S_{k_1}$, and the contribution of the cut vertices can be expressed by $\varepsilon(n_q - 1)$ with $0 \leq \varepsilon < 1$. Consequently, the total contribution of $\sum C_i$ can be expressed by $n - (n_q - 1)(1 - \varepsilon) - S_{k_1}$, which implies values of $C_{ws}$ that are smaller than those for the structure of 'networks of cliques' arranged in a star layout, but may be close depending on the contribution of the connecting vertices.

For the clustering coefficient (transitivity) $C$ given by Eq. (16), we can find expressions for 'networks of cliques'. By the definition of clique, each one of them contributes with $\binom{q_i}{3}$ triangles, where $q_i$ is the size of the $i$th clique. The total number of triangles of 'networks of cliques' that are minimally connected (and theoretically for the initial state of isolated cliques) is given by Eq. (18).

$$N_{\triangle_0} = \frac{1}{6}\sum_{i=1}^{n_q} q_i(q_i - 1)\,(q_i - 2). \tag{18}$$

The calculation of the number of triples requires the identification of vertices of two categories: (i) vertices of the cliques and (ii) connection vertices, which are also counted in the first category. Each vertex of a clique contributes with $k_i(k_i - 1)/2$ triples, where $k_i$ is the vertex degree. As the degrees of the vertices of a clique are equal to $(q_i - 1)$, we can express the number of triples of each vertex by $(q_i - 1)(q_i - 2)/2$. The total number of triples for each clique is $q_i(q_i - 1)(q_i - 2)/2$.

For each connection vertex $v_j$, the number of triples is given by $k_j \cdot k_k$, where $k_j$ and $k_k$ represent the degrees of two adjacent cliques.

Each cut point or cut vertex (i.e., the white vertices in Fig. 2(a), (d) and (c)) or connection vertex (i.e., the white vertices in Fig. 2(b)) contributes with $(q_j - 1)(q_k - 1)$ triples. From the above, the total number of connected triples of vertices is given by Eq. (19).

$$N_T = \frac{1}{2}\sum_{i=1}^{n_q} q_i(q_i-1)\,(q_i-2) + \sum (q_j-1)\,(q_k-1). \tag{19}$$

Substituting these quantities into Eq. (16), the transitivity for the theoretical structures of 'networks of cliques' that are minimally connected is given by Eq. (20).

$$C_q = \frac{3N_{\Delta_0}}{N_T} = \frac{\frac{1}{2}\sum_{i=1}^{n_q} q_i(q_i-1)\,(q_i-2)}{\frac{1}{2}\sum_{i=1}^{n_q} q_i(q_i-1)\,(q_i-2) + \sum (q_j-1)\,(q_k-1)}. \tag{20}$$

Alternatively, to simplify the interpretation of the coefficient, we can rewrite Eq. (20), as follows:

$$\frac{1}{C_q} = 1 + \frac{\sum (q_j-1)\,(q_k-1)}{\frac{1}{2}\sum_{i=1}^{n_q} q_i(q_i-1)\,(q_i-2)}. \tag{21}$$

When a 'network of cliques' has only one clique, there are no cut or connection vertices; then, the numerator ($\sum (q_j-1)\,(q_k-1)$) of Eq. (21) is equal to 0 and $C_q = 1$. Another way to reach the same conclusion is to note that, in this case, $q_j = q_k$ for every vertex. It is also important to observe that the transitivity for the structures of 'networks of cliques' that are minimally connected depends not only on the theoretical structure but also on how the cliques are connected according to their sizes. This scenario occurs because the number of transitive triples of the cut or connection vertices varies with the cliques' arrangement in the network. However, for all of the theoretical structures, the basic principle is that the transitivity is minimal if the product of the sizes of the adjacent cliques $(q_j-1)(q_k-1)$ is at a maximum. In addition, the transitivity can be calculated by properly arranging the cliques by their sizes, maximizing these products.[1]

Based on the foregoing, the 'networks of cliques' arranged in a star layout is the arrangement that has the lowest transitivity compared to networks with the same amount of cliques of corresponding sizes (i.e., the same distribution of clique sizes).

This scenario occurs because, when calculating the contribution of the triplets that have the cut points, there are $\binom{n_q}{2} = n_q(n_q-1)/2$ pairs of cliques (i.e., two cliques connected after the juxtaposition process) for the 'networks of cliques' arranged in a star layout, while for the 'networks of cliques' arranged in line and layer layouts, there are only $(n_q-1)$ pairs of cliques.[2] The number of pairs of cliques that accounts for the connection vertices of a 'network of cliques' arranged in a circle layout is $n_q$.

From the perspective of the theoretical structures of 'networks of cliques' that are minimally connected, we can summarize with the following points: (i) the structure of the 'networks of cliques' arranged in a star layout (having the maximum transitivity in terms of cliques) has $C_{ws} \simeq 1$ and $C_q \to 0$ for networks with a large number of cliques; and (ii) the structure of 'networks of cliques' arranged in a line layout (having the minimum transitivity in terms of cliques) typically has values of $C_{ws}$ that can be too high. For $C_q$, the analytical comparison for the same theoretical structure is not simple. Simulations show that these values are close to $C_{ws}$ and that $C_{ws} - C_q$ is no more than 0.12.

Therefore, a 'network of cliques' is characterized as *small-world* with respect to the transitivity between the cliques (and not only between the vertices). We expect that this type of network has $C_{ws} \simeq 1$ (which can occur for both the structures of 'networks of cliques' arranged in star and line layouts), but having $C \simeq 0$ (which occurs only for the structure of 'networks of cliques' arranged in a star layout). Otherwise, we can say that the greater the difference $C_{ws} - C$ for the real 'network of cliques' is, the more transitive the network, i.e., the *transitivity of cliques* $C_{clique} = C_{ws} - C$ quantifies the transitivity in terms of the cliques, while $C_{ws}$ and $C$ could separately have very different values for the same structure in an attempt to capture the transitivity of the network in terms of the vertices.

Using the two parameters (i.e., the reference diameter and the transitivity of the cliques) described above, we can establish the 'networks of cliques' as small-world networks that are based on their basic elements (i.e., cliques). The 'network of cliques' arranged in a star layout has the minimum diameter and the maximum transitivity of cliques $C_{clique}$. It is, therefore, reasonable that the *small-world* topology for 'networks of cliques' is one in which $C_{clique} \geq 0.5$, and the reference diameter is determined for 'networks of cliques' arranged in a star layout ($0 \leq D^*_{ref} \leq 0.25$, Table 1); this result occurs because the diameter $D$ is much closer to 2 than the maximum value $n_q$.

The quantitative summary of the proposed indices applied to the theoretical structures of 'networks of cliques' that are minimally connected (Fig. 2) is presented in Table 2.

[1] For the structure of 'networks of cliques' arranged in a star layout, the parcel $(q_j-1)\,(q_k-1)$ does not depend on the arrangement of the sizes of cliques because all of the cliques are adjacent to each other.

[2] Smaller numbers of pairs of cliques result in a higher transitivity. In the case of 'networks of cliques' arranged in line and layer layouts that have the same number of pairs of cliques, the layer structure has a lower value for the transitivity because its configuration allows most of the cliques (of size $q_r$) to pair up to $r$ other cliques. On the other hand, for the linear structure, two pairs of cliques is the maximum possible number.

**Table 2**
Quantitative summary of the indices of cohesion for the theoretical structures of 'networks of cliques' that are minimally connected (Fig. 2).

| Theoretical structures | $\Delta_{q0}$ | $\Delta$ | $\upsilon(\Delta)$ | $D$ | $n_q$ | $D^*_{ref}$ | $C_{ws}$ | $C$ | $C_{clique}$ |
|---|---|---|---|---|---|---|---|---|---|
| Line layout | 0.1080 | 0.1858 | 0.7194 | 8 | 8 | 1.0000 | 0.8093 | 0.6543 | 0.1550 |
| Circle layout | 0.1080 | 0.2035 | 0.8831 | 5 | 8 | 0.6610 | 0.7790 | 0.6474 | 0.1316 |
| Star layout | 0.1080 | 0.1858 | 0.7194 | 2 | 8 | 0.0000 | 0.9595 | 0.3739 | 0.5856 |
| Layer layout | 0.1080 | 0.1858 | 0.7194 | 4 | 8 | 0.5000 | 0.8429 | 0.5942 | 0.2487 |

**Table 3**
Basic statistics for the semantic networks based on the titles of the scientific papers.

| Journal | Date | Titles | Words ($n_0$) | Vertices ($n$) | $\lvert\mathcal{E}_0\rvert$ | $\lvert\mathcal{E}\rvert$ | $n_0/n$ | $\lvert\mathcal{E}_0\rvert/\lvert\mathcal{E}\rvert$ | Comp. |
|---|---|---|---|---|---|---|---|---|---|
| AFE | 1999 a 2008 | 371 | 3,987 | 1,561 | 20,941 | 17,880 | 2.5541 | 1.1712 | 2 |
| APPL | 1980 a 2009 | 658 | 5,113 | 1,322 | 19,228 | 14,151 | 3.8676 | 1.3588 | 1 |
| ARJG | 1969 a 2008 | 971 | 6,087 | 2,158 | 19,373 | 16,530 | 2.8207 | 1.1720 | 10 |
| CB | 1994 a 2008 | 1,642 | 13,464 | 4,225 | 56,997 | 46,560 | 3.1867 | 1.2242 | 13 |
| HR | 1947 a 2009 | 3,000 | 19,960 | 3,884 | 66,040 | 49,176 | 5.1390 | 1.3429 | 10 |
| Nature | 1997 a 2009 | 35,163 | 178,598 | 23,160 | 454,442 | 349,546 | 7.7115 | 1.3001 | 142 |
| PEM | 1986 a 2009 | 703 | 5,247 | 1,186 | 18,673 | 12,903 | 4.4241 | 1.4472 | 1 |
| PRA | 2007 a 2008 | 3,205 | 23,515 | 4,098 | 84,428 | 60,515 | 5.7382 | 1.3952 | 2 |
| PRB | 2007 a 2008 | 7,844 | 64,506 | 8,386 | 256,772 | 160,658 | 7.6921 | 1.5983 | 4 |
| PRC | 1970 a 2009 | 23,000 | 160,528 | 14,204 | 552,463 | 254,445 | 11.3016 | 2.1712 | 2 |
| PRD | 2007 a 2008 | 5,527 | 38,430 | 4,700 | 130,070 | 73,337 | 8.1766 | 1.7736 | 4 |
| PRE | 2007 a 2008 | 4,156 | 30,013 | 5,296 | 104,763 | 79,660 | 5.6671 | 1.3151 | 3 |
| PRL | 2007 a 2008 | 5,929 | 42,496 | 7,371 | 146,205 | 107,516 | 5.7653 | 1.3598 | 5 |
| Science | 1997 a 2009 | 11,192 | 74,823 | 15,240 | 233,040 | 188,346 | 4.9096 | 1.2373 | 55 |
| SHI | 1979 a 2008 | 845 | 6,431 | 2,098 | 23,943 | 20,442 | 3.0653 | 1.1713 | 1 |

## 4. Results and discussion

To study the proposed indices of cohesion, we have used a dataset that is composed of 15 scientific journals: *Agricultural and Forest Entomology*—**AFE**; *Applied Psycholinguistics: Psychological and Linguistic Studies Across Languages and Learning*—**APPL**; *Antipode: A Radical Journal of Geography*—**ARJG**; *Chemistry and Biology*—**CB**; *Human Relations: Towards the Integration of the Social Sciences*—**HR**; *Nature*—**Nature**; *Probabilistic Engineering Mechanics*—**PEM**; *Physical Review A*—**PRA**; *Physical Review B*—**PRB**; *Physical Review C*—**PRC**; *Physical Review D*—**PRD**; *Physical Review E*—**PRE**; *Physical Review Letters*—**PRL**; *Science*—**Science** and *Sociology of Health and Illness*—**SHI**.

Details of the construction of the semantic networks based on the titles of the scientific papers can be found in Refs. [10,6,11,7], and we summarize the basic statistics for the semantic networks based on the titles of the scientific papers in Table 3.

In Table 3, we observe that the networks of the journals APPL, PEM and SHI are connected. From a preliminary analysis, we also note that the number of components does not necessarily depend on the quantity of the titles (e.g., PRC, with 23,000 titles, has only 2 components; Science, with 11,192 titles, has 55 components; and Nature, with 35,163 titles, has 142 components). The vocabulary of the journals, combined with the disciplinary aspect, influences the quantity of the components.

Using the 15 semantic networks based on titles of scientific papers as the empirical dataset, classical network indices were calculated and compared with the proposed indices for 'networks of cliques'. The results are discussed as follows.

### 4.1. Density

The density, which relates the number of edges with the maximum possible in a network, is relatively low for all of the semantic networks based on the titles of the scientific papers (Table 4).

The normalized value of $\Delta$ reflects only the fact that the reference to the minimum value ($\Delta_{norm} = 0$) is associated with the initial state of isolated cliques (e.g., Fig. 1(a)) and not to a network in which the vertices are all isolated, as accounted for in a classical density calculation. Thus, the normalized density approximately scales linearly with the classical density, as shown in Fig. 4.

Moreover, the rate of density variation is an indicator of the proportion of similar vertices and edges that are absorbed for building the 'network of cliques' (i.e., the common vertices are absorbed by juxtaposition and/or overlapping processes, and the edges that connect pairs of common vertices are absorbed by the overlapping process) with respect to the initial state of isolated cliques.

For the semantic networks based on the titles of scientific papers, higher values for $\upsilon(\Delta)$ indicate that, in the process of network building, the titles are connected to each other by a greater number of words without a large reduction of edges, whereas the density variation increases with the square of the variation of vertices (Eq. (6)). It is important to note that the

**Table 4**
The density and fragmentation values of the semantic networks based on titles of scientific papers.

| Journal | Density ($\Delta$) | Norm. Dens. ($\Delta_{norm}$) | Var. Dens. $\upsilon(\Delta)$ | Frag. ($F$) | Frag. Clique ($F_{clique}$) |
|---|---|---|---|---|---|
| AFE | 0.014685 | 0.012081 | 4.572205 | 0.003839 | 0.002703 |
| APPL | 0.016206 | 0.014757 | 10.015048 | 0.000000 | 0.000000 |
| ARJG | 0.007147 | 0.006114 | 5.875532 | 0.023024 | 0.009278 |
| CB | 0.005218 | 0.004592 | 7.297165 | 0.010387 | 0.007313 |
| HR | 0.006521 | 0.006192 | 18.669641 | 0.014362 | 0.003001 |
| Nature | 0.001303 | 0.001275 | 44.742512 | 0.018032 | 0.004010 |
| PEM | 0.018362 | 0.017028 | 12.533586 | 0.000000 | 0.000000 |
| PRA | 0.007209 | 0.006905 | 22.605432 | 0.000976 | 0.000312 |
| PRB | 0.004570 | 0.004447 | 36.024910 | 0.000715 | 0.000383 |
| PRC | 0.002523 | 0.002480 | 57.829731 | 0.000141 | 0.000043 |
| PRD | 0.006641 | 0.006466 | 36.702959 | 0.002126 | 0.000543 |
| PRE | 0.005681 | 0.005450 | 23.424281 | 0.003018 | 0.000481 |
| PRL | 0.003958 | 0.003797 | 23.445694 | 0.003793 | 0.000675 |
| Science | 0.00162 | 0.001541 | 19.043115 | 0.013580 | 0.004825 |
| SHI | 0.009293 | 0.008144 | 7.024762 | 0.000000 | 0.000000 |

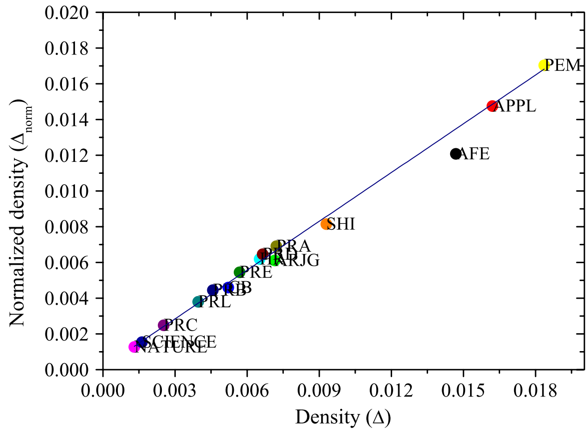



**Fig. 4.** Density $\Delta$ vs. normalized density $\Delta_{norm}$ for the semantic networks based on the titles of scientific papers. A fit was performed that was linear in $f(x) = bx$, where $b = 0.910$ with $R^2 = 0.997$.

process of building a 'network of cliques' results in high values for the density variation and reflects the predominance of the juxtaposition process or, at most, the overlapping process with three vertices, whereas in the juxtaposition process, the vertices are reduced without a reduction in the edges; in the overlapping process with two vertices, an edge is reduced, while in the overlapping process with three vertices, the reduction has the same value for the vertices and edges (from four or more vertices, the number of edges is greater than the number of vertices). This fact is shown in Fig. 5, i.e., the density variation can be interpreted by the ratio between the total number of words of the titles $n_0$ and the number of words (vertices) $n$ of the network.

In Table 4, we show that the PRC semantic network has the highest density variation. This result is related to a greater variation in the vertices among all of the networks because the variation in the edges has a slight impact on the density variation.

Furthermore, this scenario implies that the vocabulary used in the titles is more homogeneous for this network. In addition, we also show in Table 3 that the connections by the juxtaposition process are predominant; this construct can be noted in the values of the variations of the vertices ($n_0/n$) compared with the values of the variations of the edges ($|\mathcal{E}_0|/|\mathcal{E}|$).

### 4.2. Fragmentation

As mentioned above, fragmentation is an index that is used to normalize the number of components in a network. The fragmentation index $F$ (Eq. (10)) depends on the size of the components. In the 'networks of cliques' case, fragmentation depends on the relative size of the largest component, which is given by the percentage of vertices of the component in relation to the total number of vertices. The fragmentation values shown in Table 4 can best be compared by means of Fig. 6. We can note that the ARJG semantic network is more fragmented.

Some changes in the fragmentation values can also be observed (e.g., CB and HR semantic networks). For example, for the fragmentation $F$, the HR semantic network is more fragmented than the CB semantic network; on the other hand, for the fragmentation of clique $F_{clique}$, the CB semantic network is more fragmented than the HR semantic network. It follows that the components (except for the largest component) of the CB semantic network are larger than those of the HR semantic

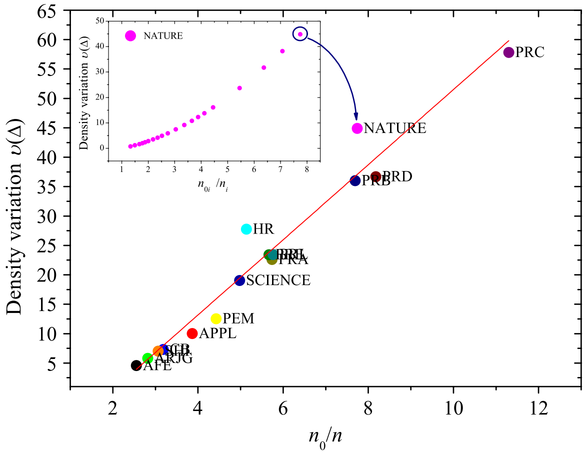



**Fig. 5.** Density variation $\upsilon(\Delta)$ vs. $n_0/n$ for the semantic networks based on the titles of scientific papers. A fit was performed that was linear in $f(x) = bx+a$, where $b = -12.367$ and $a = 6.387$ with $R^2 = 0.954$.

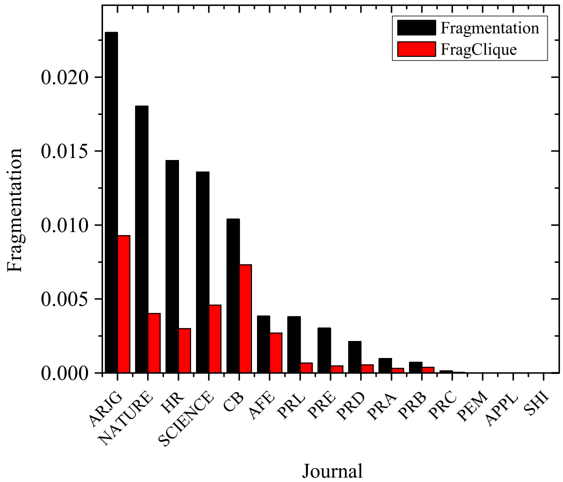



**Fig. 6.** Values of the fragmentation ($F$) and the fragmentation of cliques ($F_{clique}$) for the semantic networks based on the titles of papers.

network, which directly influences the $F$ values. For the 'network of cliques', and specifically for the semantic networks based on titles of papers, the interpretation of fragmentation must account for how the titles came together to form the network, i.e., how the titles are "defragmented" so that they are best captured by the $F_{clique}$ values. In absolute terms, the $F_{clique}$ values result from less fragmented networks (i.e., the networks are less fragmented in relation to the titles of the papers compared to in relation to the words).

### 4.3. Cut point

By definition, the indices discussed above can be applied to unconnected networks. However, other indices are difficult to interpret for unconnected networks, in spite of being subjected to calculations. Because unconnected semantic networks have a largest component with more than 99% of the vertices, it is possible to analyze and discuss appropriate indices for the connected networks, such as the cut points. For the semantic networks that are based on the titles of papers, the absolute and normalized values for the cut points are shown in Table 5.

Considering only the absolute values, the Nature semantic network has more cut points, but the normalization shows that the ARJG semantic network is a more "fragile" network. A relationship between the initial state of isolated cliques and the largest component of the specific 'network of cliques' can be made using the concepts of the cut point and the fragmentation of cliques. This relationship is shown in Fig. 7, which shows only the unconnected networks (i.e., not including APPL, PEM and SHI). The results indicate a tendency for networks with higher values for the fragmentation of cliques to have proportionately higher values for the normalized cut point index. This result means that the absence of links between the components in an unconnected network is directly reflected in its core (i.e., the largest component) because of the fragility of the connections (expressed by the normalized cut point index).

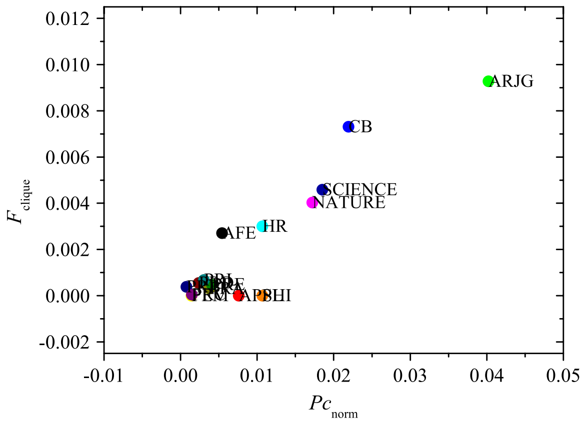



**Fig. 7.** Variation in the fragmentation of the clique vs. cut point index normalized for semantic networks based on the titles of papers. Only unconnected networks are shown.

**Table 5**
Values of the cut points ($P_c$), normalized cut points ($Pc_{norm}$, Eq. (12)), reference diameter ($D^*_{ref}$, Eq. (13)) and clustering coefficients ($C_{ws}$ and $C$, Eqs. (15) and (16), respectively; and $C_{clique} = C_{ws} - C$) of the major components of the connected semantic networks based on the titles of papers.

| Journal | $P_c$ | $Pc_{norm}$ | $D^*_{ref}$ | $C_{ws}$ | $C$ | $C_{clique}$ |
|---|---|---|---|---|---|---|
| AFE | 2 | 0.005420 | 0.1755 | 0.7919 | 0.2079 | 0.5840 |
| APPL | 5 | 0.007610 | 0.1581 | 0.7578 | 0.1679 | 0.5899 |
| ARJG | 39 | 0.040583 | 0.2028 | 0.7717 | 0.1696 | 0.6020 |
| CB | 36 | 0.022099 | 0.1367 | 0.8058 | 0.1337 | 0.6721 |
| HR | 32 | 0.010702 | 0.1503 | 0.7346 | 0.1376 | 0.5970 |
| Nature | 602 | 0.017190 | 0.1282 | 0.6997 | 0.0832 | 0.6165 |
| PEM | 1 | 0.001425 | 0.1182 | 0.7660 | 0.1801 | 0.5860 |
| PRA | 12 | 0.003746 | 0.1242 | 0.7363 | 0.1403 | 0.5960 |
| PRB | 6 | 0.000765 | 0.0838 | 0.7881 | 0.1318 | 0.6563 |
| PRC | 34 | 0.001478 | 0.0980 | 0.8340 | 0.1080 | 0.7260 |
| PRD | 13 | 0.002354 | 0.0875 | 0.7402 | 0.1496 | 0.5907 |
| PRE | 17 | 0.004093 | 0.1200 | 0.7304 | 0.1298 | 0.6006 |
| PRL | 18 | 0.003041 | 0.1146 | 0.7510 | 0.1277 | 0.6234 |
| Science | 218 | 0.019480 | 0.1743 | 0.7285 | 0.0872 | 0.6413 |
| SHI | 9 | 0.010664 | 0.1515 | 0.7629 | 0.1543 | 0.6085 |

### 4.4. Small-world characterization

The distance and clustering coefficients are widely used to characterize a network as *small-world* [2]. However, the 'networks of cliques', by construction, usually have high values for the clustering coefficients (Eq. (15)). Especially for the semantic networks based on the titles of papers, the clustering coefficient values are between 0.69 and 0.84. These values are much higher than those for the corresponding random networks, and the geodesic distances are close to those for the random networks. According to Ref. [2], these types of networks are small-world networks.

Theoretical discussion on the proposed indices (Section 3) allows us to quantify the transitivity of the papers' titles based on the transitivity of the words, using as a reference the relationship between the concepts of the clustering coefficient $C_{ws}$ and the transitivity coefficient $C$. As the distances between the titles, these values are evaluated by the reference diameter ($D^*_{ref}$) of the network. Reference diameters near to 0 (zero) are compatible with the 'networks of cliques' arranged in a star layout, and at the other end, reference diameters close to 1 (one) are compatible with the 'networks of cliques' arranged in a line layout. In Table 5, we show that the reference diameters for the semantic networks based on the titles of papers are in the range of the 'networks of cliques' arranged in a star layout (Table 1). On the other hand, the relationship between the clustering and transitivity coefficients satisfies the recommendations, because all semantic networks have $C_{clique} > 0.5$, as shown in Table 5. These results leave no doubt about the small-world topology presented by the semantic networks based on the titles of papers in relation to the words of the titles, as observed in Refs. [6,7]. Moreover, the results show that a small-world topology occurs in relation to the titles (i.e., the cliques). In sum, the topological characterization is intended to cover not only the formation of clusters of vertices, which is trivial in network of cliques, but also the clusters of cliques, which is less trivial.

## 5. Concluding remarks

These research results allow us to show that some of the classical parameters used in the analysis of networks are not entirely suitable for 'networks of cliques'. The traditional analysis does not take into account the juxtaposition and overlapping processes in the formation of 'network of cliques'. A network of cliques, as defined in this paper, has particular characteristics: some indices have high values in comparing to non-network of cliques (e.g. density, degree of vertex), due to the initial state of isolated cliques. Within this context, we can measure and interpret more appropriately the studied networks.

Density variation ($\upsilon(\Delta)$) allows us to infer with respect to the homogeneity of the vocabulary of the semantic network, and it is an index that is directly related to the quotient between the number of vertices of the initial state of isolated cliques and the number of vertices of the 'networks of cliques' after the processes of juxtaposition and/or the overlapping of cliques. For semantic networks that are based on titles of papers, the density variation scales linearly with the quotient word/vertex because this variation was found to be predominant over the variation of edges.

The fragmentation of cliques and the cut point index are more suitable for 'networks of cliques' and exhibit a close relationship. In addition, for the semantic networks based on the titles of papers, these indices linearly relate the number of unconnected titles to the fragility of the largest component on the withdrawal of vertices.

Finally, the characterization of a semantic network based on the titles of papers (such as small-world, via the geodesic distance and the clustering coefficient of words) becomes somewhat trivial by virtue of the construction of semantic networks (i.e., 'networks of cliques'). The parameters used in this research (i.e., the diameter and transitivity coefficient of the cliques) also enable us to investigate the small-world feature from the titles' perspective. This study opens a fruitful field for further investigations, such as the degree distribution and dynamic behavior of 'networks of cliques', which is not addressed here.

## References

[1] J. Gross, J. Yellen, Graph Theory and its Applications, CRC Press, Boca Raton, 1999.
[2] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440.
[3] A.-L. Barabási, R. Albert, Science 286 (1999) 509.
[4] L.A.N. Amaral, A. Scala, M. Barthélémy, H.E. Stanley, Proc. Natl. Acad. Sci. 97 (2000) 11149.
[5] M.E.J. Newman, SIAM Rev. 45 (2003) 167.
[6] I.S. Fadigas, T. Henrique, V. de Senna, M.A. Moret, H.B.B. Pereira, Educação Matemática Pesquisa 11 (2009) 167.
[7] H.B.B. Pereira, I.S. Fadigas, V. Senna, M.A. de and Moret, Physica A 390 (2011) 1192.
[8] S.P. Borgatti, Comput. Math. Organiz. Theor. 12 (2006) 21.
[9] M.E.J. Newman, S.H. Strogatz, D.J. Watts, Phys. Rev. E 64 (2001) 026118.
[10] S.M.G. Caldeira, T.C.P. Lobão, R.F.S. Andrade, A. Neme, J.G.V. Miranda, Eur. Phys. J. B 49 (2006) 523.
[11] G.M. Teixeira, M.S.F. Aguiar, C.F. Carvalho, D.R. Dantas, M.V. Cunha, J.H.M. Morais, H.B.B. Pereira, J.G.V. Miranda, Int. J. Mod. Phys. C 21 (2010) 333.