When local and global clustering of networks diverge

Ernesto Estrada

Department of Mathematics & Statistics, University of Strathclyde, 26 Richmond Street, Glasgow G1 1HX, UK

A R T I C L E   I N F O

A B S T R A C T

The average Watts–Strogatz clustering coefficient and the network transitivity are widely used descriptors for characterizing the transitivity of relations in real-world graphs (networks). These indices are bounded between zero and one, with low values indicating poor transitivity and large ones indicating a high proportion of closed triads in the graphs. Here, we prove that these two indices diverge for windmill graphs when the number of nodes tends to infinity. We also give evidence that this divergence occurs in many real-world networks, especially in citation and collaboration graphs. We obtain analytic expressions for the eigenvalues and eigenvectors of the adjacency and the Laplacian matrices of the windmill graphs. Using this information we show the main characteristics of two dynamical processes when taking place on windmill graphs: synchronization and epidemic spreading. Finally, we show that many of the structural and dynamical properties of a real-world citation network are well reproduced by the appropriate windmill graph, showing the potential of these graphs as models for certain classes of real-world networks.

E-mail address: ernesto.estrada@strath.ac.uk.

1. Introduction

The development of network theory has demanded the definition of several mathematical indices that characterize certain aspects of the topology of these usually giant graphs [1–3]. In a seminal paper published in 1998 Watts and Strogatz [4] defined an index nowadays known as the *Watts–Strogatz clustering coefficient* of a given node in a graph. This index accounts for the ratio of the number of triangles in which the corresponding node takes place to the number of potential triangles involving that node (see further for a formal definition). The clustering coefficient is bounded between zero and one. An index close to zero indicates that the relative number of transitive relations involving that node is low. A clustering coefficient close to one indicates that this node is involved in as many transitive relations as possible. When studying complex real-world networks it is very frequent to report the average Watts–Strogatz (WS) clustering coefficient as a characterization of how globally clustered a network is. Bollobás [5] has remarked that: "*this kind of 'average of an average' is often not very informative*".

An alternative way of characterizing the transitivity of a network is by taking the total number of triangles in the graph divided by the total number of triads existing in the graph. This index was put forward originally by Luce and Perry [6] and then rediscovered by Newman [7] in the context of complex networks. It is commonly referred to as the *transitivity index* of a (social) network (see for instance [8]). It is straightforward to realize that the transitivity index is also bounded between zero and one, with small values indicating poor transitivity and values close to one indicating a large one.

Due to the previous definitions of the WS average clustering coefficient and the transitivity index it is very surprising that there are simple graphs for which both indices diverge. In the Fig. 1 we give a couple of examples for which the WS average clustering coefficient tends asymptotically to zero while the transitivity index goes to one when the size of the graph tends to infinity. The first graph is known as the *friendship* (or Dutch windmill or *n*-fan) *graph* and the example was reported by the first time by Bollobás [5]. The second example was reported firstly by Estrada in [1] and these graphs were further named *agave graphs* by Estrada and Estrada-Vargas in a different context [9]. According to these examples, for a relatively large number of nodes, the WS average clustering coefficient tells us that the graphs are as transitive as possible, while the transitivity index is telling us exactly the contrary.
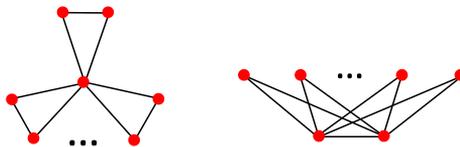
Fig. 1. Two examples of classes of graphs for which the Watts–Strogatz clustering coefficient and the graph transitivity diverge when the number of nodes tends to infinity.
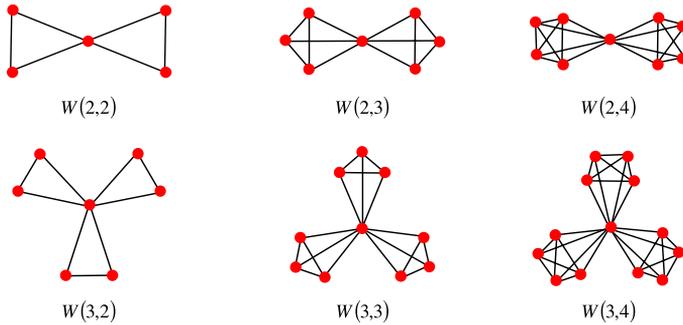
Fig. 2. Illustration of the smallest windmill graphs $W(\eta, \kappa)$ consisting of η copies of the complete graph K_κ.

The friendship graphs are formed by gluing together κ copies of a triangle in a common vertex. These graphs appeared as the solution of the *Friendship Theorem* of Erdős, Rényi and Sós [10], which states that the only graphs with the property that every two vertices have exactly one neighbor in common are the friendship graphs. The friendship graphs are members of a larger family of graphs known as the *windmill graphs*. A windmill graph $W(\eta, \kappa)$ consists of η copies of the complete graph K_κ [11] with every node connected to a common one (see Fig. 2). We remind that a complete graph K_κ is the graph with κ nodes and $\kappa(\kappa - 1)/2$ edges. Thus, the friendship graphs are just $W(\eta, 2)$.

Although the friendship graphs have received recent attention due to their spectral properties [12–15], the general class of windmill graphs has received little attention apart from their study in the topic of graceful and harmonious graph labeling (see [16] for a review). Here, we study this family of graphs in relation to the problem of clustering divergence, which is defined as the divergence of the average WS and the transitivity index when the graph size tends to infinity. We study the spectra of the adjacency and the Laplacian matrices of these graphs and using these results we analyze the particularities of some dynamical processes taking place on them. We conclude that windmill graphs are bad structures for seeking synchronization. In addition, an epidemic dynamics running on them shows null epidemic threshold. We then show that the clustering divergence phenomenon is also observed in many real-world networks, in particular in citation and in collaboration networks (something unnoticed so far in the literature). We show that the windmill graphs are appropriated cores for building models capturing the clustering divergence in real-world networks as well as for reproducing the structural and dynamical properties of these real-world graphs.

2. Preliminaries

A *graph* $G = (V, E)$ is defined by a set of n nodes (vertices) V and a set of m edges $E = \{(u, v) | u, v \in V\}$ between the nodes. The *degree* of a vertex, denoted by k_i, is the number of edges incident to i in G. A graph with unweighted edges, no edges from a node to itself, and no multiple edges is said to be *simple*. We will always consider undirected,

simple, and connected networks. The so-called *Watts–Strogatz clustering coefficient* of a node i, which quantifies the degree of transitivity of local relations in a network is defined as [4]:

$$C_i = \frac{2t_i}{k_i(k_i - 1)} \tag{1}$$

where t_i is the number of triangles in which the node i takes place. Taking the mean of these values as i varies among all the nodes in G, one gets the *average WS clustering coefficient* of the network G:

$$\overline{C} = \frac{1}{n} \sum_{u=1}^{n} C_u. \tag{2}$$

The so-called graph transitivity is defined as [7,8]

$$C = \frac{3t}{P_2}, \tag{3}$$

where t is the total number of triangles and $P_2 = \sum_{i=1}^{n} k_i(k_i - 1)/2$.

Let $A = (a_{uv})_{n \times n}$ be the *adjacency matrix* of the graph. We denote by $\lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n$ the eigenvalues of A and by $\psi_{j,p}$ the pth entry of the jth orthonormalized eigenvector $\vec{\psi}_j$ associated with λ_j. For the whole spectrum of a given matrix M we use the following notation: $Sp(M) = \{\lambda^m\}$ where λ is an eigenvalue and m its algebraic multiplicity.

The adjacency matrix of a windmill graph $W(\eta, \kappa)$ can be written as

$$A(W(\eta, \kappa)) = \begin{pmatrix} 0 & \vec{1}^T & \vec{1}^T & \cdots & \vec{1}^T \\ \vec{1} & A(K_\kappa) & 0 & \cdots & 0 \\ \vec{1} & 0 & A(K_\kappa) & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vec{1} & 0 & 0 & \cdots & A(K_\kappa) \end{pmatrix}, \tag{4}$$

where $\vec{1}$ is an all-ones column vector of length κ and $A(K_\kappa)$ is the adjacency matrix of K_κ.

Let Δ be a diagonal matrix of the vertex degrees of the graph. Then, the Laplacian matrix of the graph is defined by: $L = \Delta - A$. In the case of the windmill graphs it can be written as

$$L(W(\eta, \kappa)) = \begin{pmatrix} \eta\kappa & -\vec{1}^T & -\vec{1}^T & \cdots & -\vec{1}^T \\ -\vec{1} & L(K_\kappa) + I_\kappa & 0 & \cdots & 0 \\ -\vec{1} & 0 & L(K_\kappa) + I_\kappa & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\vec{1} & 0 & 0 & \cdots & L(K_\kappa) + I_\kappa \end{pmatrix}, \tag{5}$$

where I_κ is the $\kappa \times \kappa$ identity matrix. We denote by $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n = 0$ the eigenvalues of L. The eigenvalue μ_{n-1} is known as the *algebraic connectivity* of the graph and the eigenvector associated to it is known as the *Fiedler vector [17]*. For reviews about spectral properties of graphs the reader is directed to the classic works [18,19].

3. Clustering of windmill graphs

We start here by proving that the average Watts–Strogatz coefficient and the transitivity index diverge when the number of nodes increases in a windmill graph.

Theorem 1. *Let $W(\eta, \kappa)$ be a windmill graph. Then, for a given value of κ the average Watts–Strogatz and the transitivity coefficients diverge when the number of cliques tends to infinity:*

$$\lim_{\eta \to \infty} \bar{C} = 1, \tag{6}$$

$$\lim_{\eta \to \infty} C = 0. \tag{7}$$

Proof. First, we obtain an expression for the Watts–Strogatz clustering coefficient of windmill graphs. The Watts–Strogatz clustering coefficient of the central node in $W(\eta, \kappa)$ is

$$C_i = \frac{\kappa - 1}{\eta\kappa - 1}, \tag{8}$$

and the rest of the nodes have $C_j = 1$. Thus, $\bar{C} = (C_i + \eta\kappa C_j)/n$, which gives

$$\bar{C} = 1 - \frac{\kappa(\eta + 1)}{(\eta\kappa + 1)(\eta\kappa - 1)}. \tag{9}$$

Now we consider the transitivity index of a windmill graph $W(\eta, \kappa)$. The total number of triangles in $W(\eta, \kappa)$ is
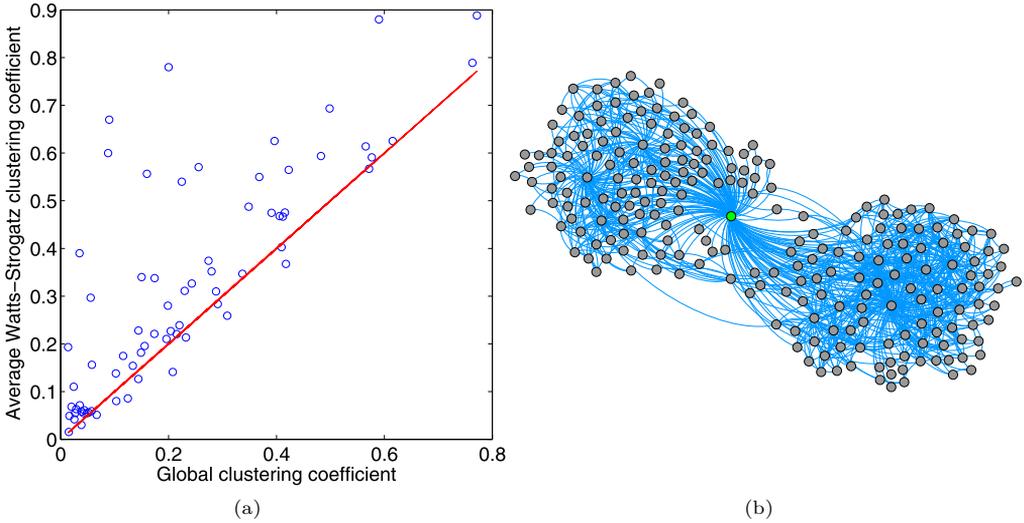
$$t = \frac{1}{6}\eta\kappa\left(\kappa^2 - 1\right). \tag{10}$$

In order to obtain the number of 2-paths we use

$$P_2 = \frac{1}{2}k_i(k_i - 1) + \frac{\eta\kappa}{2}k_j(k_j - 1) \tag{11}$$

in which $k_i = \eta\kappa$ and $k_j = \kappa$. Thus, after substitution we get

$$C = \frac{\kappa^2 - 1}{\kappa^2 + \eta\kappa - \kappa - 1}. \tag{12}$$

Fig. 3. (a) Scatterplot of the average Watts–Strogatz clustering coefficient $\bar{C}$ and the global transitivity C for a series of real-world networks. The red line corresponds to $\bar{C} = C$. (b) Illustration of the citation graph in the field of Small World networks, which has $\bar{C} \approx 0.556$ and $C \approx 0.160$. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Obviously, $\lim_{\eta \to \infty} \bar{C} = 1$ and $\lim_{\eta \to \infty} C = 0$ for a given value of κ, which proves the result. $\square$

This result clearly points out to the fact that the structural information accounted for by the average Watts–Strogatz coefficient and the transitivity index are of different nature. Clearly, when $n \to \infty$ in a windmill graph, the number of nodes with clustering coefficient equal to one grows to infinity. However, the graph as a whole remains poorly clustered due to the fact that there are no triangles involving more than one cluster K_κ. That is, a *windmill graph is locally clustered but globally poorly clustered*. This is exactly what the two clustering coefficients reflect for these graphs. Then, it is easy to see that *the average Watts–Strogatz clustering coefficient $\bar{C}$ quantifies how much, on average, a graph is locally clustered, while C accounts for the clustering of the network as a whole.*

The situation previously described is not exclusive of windmill graphs. That is, the existence of graphs with large averaged local clustering and very small global transitivity appears to be present also in real-world networks. For instance, we have investigated the relation between the average Watts–Strogatz and the global transitivity coefficients for 72 real-world networks describing complex systems in a variety of scenarios ranging from social and ecological to infrastructural and biological ones. This dataset of networks is formed by 66 real-world networks described in the Appendix of [1] and 6 ones described in [2]. For a full description of these networks the reader is referred to these two references. In the Fig. 3 (a) we illustrate the plot of the two clustering coefficients, where the points close to the top-left corner correspond to networks with large local and small global clustering. For instance, the network of film actors in which everyone of the 449,913

nodes represents an actor and each of the 25,516,482 edges represent a pair of actors that have acted in the same film has $\bar{C} \approx 0.78$ and $C \approx 0.20$ [2]. Similarly, the network of collaboration among scientists working in the area of biology, which has 1,520,251 nodes and 11,803,064 edges, displays similar characteristics with $\bar{C} \approx 0.60$ and $C \approx 0.088$ [2]. In the Fig. 3 (b) we illustrate the citation network of papers citing Milgram's 1967 Psychology Today paper or use "Small World" in the title (see [1] for details). The nodes represent papers and two nodes are connected if one paper cites the other (we consider the undirected version of this network only). It consists of 233 papers and 994 citations. This network has $\bar{C} \approx 0.556$ and $C \approx 0.160$. The node marked in a different color in the Fig. 3 (b) corresponds to Milgram paper and it is clear that at least two big groups of nodes exist in the network (see further analysis).

4. Spectral analysis of regular windmill graphs

Due to the importance that spectral properties of graphs have on a variety of structural and dynamical properties of graphs we embark here in the analysis of the spectra of windmill graphs. We start by analyzing the spectra of the adjacency matrix of these graphs which is reported using the notation fixed in the previous section.

Theorem 2. *The spectrum of the adjacency matrix of a windmill graph $W(\eta, \kappa)$ is*

$$Sp\left(A\left(W\left(\eta, \kappa\right)\right)\right) = \left\{ \left(\frac{\kappa - 1}{2} \pm \sqrt{\left(\frac{\kappa - 1}{2}\right)^2 + \eta\kappa} \right), (\kappa - 1)^{\eta - 1}, -1^{\eta(\kappa - 1)} \right\}. \quad (13)$$

Proof. Let us consider the vector $\vec{\gamma} = \begin{bmatrix} x, & \dfrac{1 - x}{\eta\kappa}, & \cdots, & \dfrac{1 - x}{\eta\kappa} \end{bmatrix}^T$ for some $x < 1$. Then,

$$A\left(W\left(\eta, \kappa\right)\right) \begin{pmatrix} x \\ \dfrac{1 - x}{\eta\kappa} \\ \vdots \\ \dfrac{1 - x}{\eta\kappa} \end{pmatrix} = \begin{pmatrix} 1 - x \\ x + (\kappa - 1)\left(\dfrac{1 - x}{\eta\kappa}\right) \\ \vdots \\ x + (\kappa - 1)\left(\dfrac{1 - x}{\eta\kappa}\right) \end{pmatrix}. \quad (14)$$

By equalizing the right-hand side of (2) to $\lambda\vec{\gamma}$ we obtain

$$1 - x = \lambda x, \quad (15)$$

$$x + (\kappa - 1)\left(\frac{1 - x}{\eta\kappa}\right) = \lambda\left(\frac{1 - x}{\eta\kappa}\right), \quad (16)$$

which has solution if

$$\lambda^2 - \lambda(\kappa - 1) - \eta\kappa = 0. \quad (17)$$

Thus,

$$\lambda = \frac{\kappa - 1}{2} \pm \sqrt{\left(\frac{\kappa - 1}{2}\right)^2 + \eta\kappa}. \tag{18}$$

Thus, the two values of (18) correspond to two eigenvalues of $A\left(W\left(\eta, \kappa\right)\right)$ both with multiplicity 1.

Let $\vec{\vartheta} = \begin{bmatrix} 0 & \alpha_{1_1} & \cdots & \alpha_{1_\kappa} & \alpha_{2_1} & \cdots & \alpha_{2_\kappa} & \cdots & \alpha_{\eta_1} & \cdots & \alpha_{\eta_\kappa} \end{bmatrix}^T$ be a vector such that $\alpha_{k_j} = \alpha_k$ for all k, $\alpha_1 + \alpha_2 + \cdots + \alpha_\eta = 0$, and $\alpha_k \neq 0$ for some k. Then, it is easy to check that

$$A\vec{\vartheta} = \left(\kappa - 1\right)\vec{\vartheta}. \tag{19}$$

There are $\eta - 1$ orthogonal vectors fulfilling the conditions that $\vec{\vartheta}$ has. Thus, $\kappa - 1$ is an eigenvalue of $A\left(W\left(\eta, \kappa\right)\right)$ with multiplicity $\eta - 1$.

Now, let $\vec{\nu} = \begin{bmatrix} 0 & \alpha_{1_1} & \cdots & \alpha_{1_\kappa} & \alpha_{2_1} & \cdots & \alpha_{2_\kappa} & \cdots & \alpha_{\eta_1} & \cdots & \alpha_{\eta_\kappa} \end{bmatrix}^T$ be a vector such that $\sum_{j=1}^{\kappa} \alpha_{k_j} = 0$ for all k, and $\alpha_k \neq 0$ for some k. Then,

$$A\vec{\nu} = -\vec{\nu}. \tag{20}$$

There are $\eta\left(\kappa - 1\right)$ orthogonal vectors $\vec{\nu}$ and so -1 is an eigenvalue of $A\left(W\left(\eta, \kappa\right)\right)$ with multiplicity $\eta\left(\kappa - 1\right)$. So far we have found $\left(\eta - 1\right) + \eta\left(\kappa - 1\right) + 2$ eigenvalues, which is the total number of eigenvalues existing for $A\left(W\left(\eta, \kappa\right)\right)$, i.e., $n = \eta\kappa + 1$, which proves the result. □

Remarks.

- The characteristic polynomial of $A\left(W\left(\eta, \kappa\right)\right)$ is

$$P\left(\lambda\right) = \left(\lambda + 1\right)^{\eta(\kappa - 1)}\left(\lambda - \kappa + 1\right)^{\eta - 1}\left(\lambda^2 - \left(\kappa - 1\right)\lambda - \left(n - 1\right)\right). \tag{21}$$

- The number of triangles in $W\left(\eta, \kappa\right)$ is given by

$$t = \frac{1}{6}\sum_{j=1}^{n}\lambda_j^3 = \frac{1}{6}\eta\kappa\left(\kappa^2 - 1\right). \tag{22}$$

- The eigenvector centrality (see for instance [1,3] for definition and uses of centrality measures) of a windmill graph $W\left(\eta, \kappa\right)$ is a vector proportional to:

$$\vec{\gamma} = \begin{bmatrix} x, & \dfrac{1 - x}{\eta\kappa}, & \cdots, & \dfrac{1 - x}{\eta\kappa} \end{bmatrix}^T. \tag{23}$$

We now turn our analysis to the spectrum of the Laplacian matrix of windmill graphs. The Laplacian matrix, which is the discrete analogous of the Laplace–Bertrami operator, plays a fundamental role in the analysis of dynamical processes on networks.

Theorem 3. *The Laplacian spectrum of a windmill graph $W(\eta, \kappa)$ is*

$$Sp\left(L\left(W\left(\eta, \kappa\right)\right)\right) = \left\{\left(\eta\kappa + 1\right), \left(\kappa + 1\right)^{\eta(\kappa-1)}, 1^{\eta-1}, 0\right\}. \tag{24}$$

Proof. The fact that the Laplacian matrix has a zero eigenvalue is well-known and easy to prove—just post-multiply L by the vector $\vec{1}$—and because the windmill graph is connected the multiplicity of this eigenvalue is one. Now, let us consider the vector $\vec{\gamma} = \left[\eta\kappa y, \quad -y, \quad \cdots, \quad -y\right]^T$ for some $y > 0$. Then,

$$L\left(W\left(\eta, \kappa\right)\right) \begin{pmatrix} \eta\kappa y \\ -y \\ \vdots \\ -y \end{pmatrix} = \begin{pmatrix} \left(\left(\eta\kappa\right)^2 + \eta\kappa\right) y \\ -y\left(\eta\kappa + 1\right) \\ \vdots \\ -y\left(\eta\kappa + 1\right) \end{pmatrix} = \left(\eta\kappa + 1\right) \begin{pmatrix} \eta\kappa y \\ -y \\ \vdots \\ -y \end{pmatrix}, \tag{25}$$

proving that $n = \eta\kappa + 1$ is an eigenvalue of $L\left(W\left(\eta, \kappa\right)\right)$ with multiplicity one.

Using similar techniques as for the Theorem 2 we have the following. Let $\vec{\vartheta} = \left[0 \quad \alpha_{1_1} \quad \cdots \quad \alpha_{1_\kappa} \quad \alpha_{2_1} \quad \cdots \quad \alpha_{2_\kappa} \quad \cdots \quad \alpha_{\eta_1} \quad \cdots \quad \alpha_{\eta_\kappa}\right]^T$ be a vector such that $\alpha_{k_j} = \alpha_k$ for all k, $\alpha_1 + \alpha_2 + \cdots + \alpha_\eta = 0$, and $\alpha_k \neq 0$ for some k. Then, it is easy to check that

$$L\vec{\vartheta} = \vec{\vartheta}. \tag{26}$$

Because, there are $\eta - 1$ orthogonal $\vec{\vartheta}$ vectors, one is an eigenvalue of $L\left(W\left(\eta, \kappa\right)\right)$ with multiplicity $\eta - 1$.

Now, let $\vec{v} = \left[0 \quad \alpha_{1_1} \quad \cdots \quad \alpha_{1_\kappa} \quad \alpha_{2_1} \quad \cdots \quad \alpha_{2_\kappa} \quad \cdots \quad \alpha_{\eta_1} \quad \cdots \quad \alpha_{\eta_\kappa}\right]^T$ be a vector such that $\sum_{j=1}^{\kappa} \alpha_{k_j} = 0$ for all k, and $\alpha_k \neq 0$ for some k. Then,

$$L\vec{v} = \left(\kappa + 1\right)\vec{v}. \tag{27}$$

There are $\eta\left(\kappa - 1\right)$ orthogonal vectors $\vec{v}$ and so $\kappa + 1$ is an eigenvalue of $L\left(W\left(\eta, \kappa\right)\right)$ with multiplicity $\eta\left(\kappa - 1\right)$, which finally proves the result. $\square$

Remarks.

- The algebraic connectivity of a windmill graph $W\left(\eta, \kappa\right)$ is 1.
- The Fiedler vector of a windmill graph $W\left(\eta, \kappa\right)$ is a vector proportional to:

$$\vec{\vartheta} = \left[0 \quad \alpha_{1_1} \quad \cdots \quad \alpha_{1_\kappa} \quad \alpha_{2_1} \quad \cdots \quad \alpha_{2_\kappa} \quad \cdots \quad \alpha_{\eta_1} \quad \cdots \quad \alpha_{\eta_\kappa}\right]^T, \tag{28}$$

such that $\alpha_{k_j} = \alpha_k$ for all k, $\alpha_1 + \alpha_2 + \cdots + \alpha_\eta = 0$, and $\alpha_k \neq 0$ for some k.

5. Dynamical processes on windmill graphs

There are many different dynamical processes that can be studied on the nodes and edges of graphs. The reader is referred to [20,21] for a detailed account of the topic. One of the most important dynamical processes studied on graphs is the epidemic dynamics. The spreading of an infectious disease on networks can be modeled representing individuals as nodes and the contacts between them as edges. In this context individuals are categorized in different compartments according to their health state [22]: i.e., *susceptible* (S) for individuals that can be infected by the disease, *infected* (I) for infectious individuals that can spread the pathogen or *recovered* (R) for individuals that already passed the disease and are immune to it. Some of the classical models used in this context are the so-called SIS and SIR [22,23]. The spreading of a disease is driven in these models by the ratio among the probability β for a susceptible individual to catch the disease once in contact with an infected one through a link of the network, and the recovery rate ρ that characterizes the rate at which an infected individual recovers from the disease. Depending on its infectious power two distinct phases are possible: an absorbing one where the spreading is not efficient enough to reach a large fraction of the system and the disease is absorbed and an active phase where the epidemics reaches a macroscopic fraction of the network. The transition from the absorbing to the active phase strictly resembles a non-equilibrium second order phase transition in statistical physics [24]. The critical value of this transition $\left(\frac{\beta}{\rho}\right)_c = \tau$ is defined as the *epidemic threshold*, which in the case of an epidemic spreading on a graph has been shown to be [25–27]:

$$\tau = \frac{1}{\lambda_1}. \tag{29}$$

In the case of a windmill graph the epidemic dynamics can be an appropriate tool to understand for instance the propagation of some ideas through the citation of papers which become viral very quickly. In this case it is straightforward to realize that the epidemic threshold is

$$\tau = \frac{2}{\kappa - 1 + \sqrt{(\kappa - 1)^2 + 4\eta\kappa}}. \tag{30}$$

The immediate consequence of this relation is that for a given value of κ the epidemic threshold vanishes as $\eta \to \infty$. In other words, in a sufficiently large windmill graph there is no resistance to the propagation of an epidemic. An idea can become viral in the whole network almost without any resistance. Even worse, when κ is sufficiently large the epidemic threshold is almost null even for very small values of η. This situation resembles the one that has been largely studied for random networks with power-law degree distributions. In those models, like the Barabási–Albert one, it has been proved analytically that the epidemic threshold vanishes for very large graphs [28]. The main

cause of this effect in the so-called scale-free graphs is the existence of an uneven degree distribution with a few hubs concentrating a large number of connections. In the windmill graphs the situation is quite similar due to the existence of a node which is connected to everyone else in the network. Thus, once this node is infected it immediately can infect the rest of the network in just one step.

As another important example of a dynamical process on graph we consider now the synchronization process on networks [29,30]. Network synchronization finds many important applications in the natural and physical worlds, such as in human learning [31] and in wireless sensor networks [32]. In the particular case of windmill graphs this process is appropriate to analyze how easily a group of collaborators forming such kind of structures reaches consensus about a given discussion topic.

A large class of dynamic networks have a bounded synchronized region specified by

$$c\mu_{n-1}/\mu_1 \in (\alpha_2, \alpha_3) \subset (0, \infty), \tag{31}$$

where constants α_2, α_3 depend only on the node dynamics, and a bigger eigenratio μ_{n-1}/μ_1 implies a better network synchronizability, which likewise means a smaller coupling strength is needed [33]. Thus, the following *sync-index* is adopted:

$$Q = \frac{\mu_{n-1}}{\mu_1}, \tag{32}$$

which for the case of windmill graphs becomes

$$Q = \frac{1}{\eta\kappa + 1} = \frac{1}{n}. \tag{33}$$

As stated before, for good network synchronizability it is necessary that the sync-index is as large as possible. Then, it is easy to see that because when $n \to \infty$ the eigenratio $Q \to 0$, which means that the windmill graphs are very bad synchronizers. In other words, although synchronization among all the nodes can be reached relatively easy, the stability of the sync state is very low.

6. Windmill-like structure in real-world networks

In a previous section we have seen that there are real-world networks for which the local and the global clustering coefficients diverge. Clearly, the windmill graphs can be used as a model to describe this characteristic of the previously mentioned networks. Windmill-like structures are plausible in certain kinds of real-world situations in which there are many groups of tightly connected nodes, all of them tied to a central node. This could be the case of citations and collaboration networks. In the first case, a seminal paper could be cited by many different groups working in related but different fields. For instance, the seminal papers of Milgram in 1967 [34], Watts and Strogatz in 1998 [4] and

Table 1
Topological properties of the real-world network representing a citation network in the topic of "Small World Networks" and four different models. The models included are the windmill graph $W(33, 7)$, Erdős–Rényi (ER) $G_{ER}(233, 994)$, Barabási–Albert (BA) $G_{BA}(233, 4)$, Watts–Strogatz (WS) $G_{BA}(233, 8, 0.05)$ and random geometric graph (RRG) $G_{RRG}(233, 0.115)$.

	SmallW	$W(33, 7)$	ER	BA	WS	RGG
n	233	232	233	233	233	233
m	994	924	994	870	932	986
$\bar{C}$	0.556	0.996	0.038	0.127	**0.565**	0.617
C	0.160	**0.176**	0.037	0.065	0.559	0.592
$\bar{C}/C$	3.475	**5.659**	1.027	1.954	1.011	1.042
$\bar{l}$	2.371	1.966	2.759	**2.637**	4.353	6.710
D	4	2	5	**4**	8	16
k_{max}	147	**231**	17	84	10	18
Cl	7	**8**	3	4	5	9
λ_1	20.96	**18.49**	9.57	13.842	8.070	12.736
Q	0.0043	**0.0043**	0.0682	0.0230	0.0154	0.0029

Barabási and Albert in 1999 [35] are frequently cited by groups working in mathematics, physics, social sciences, computer sciences and neurosciences (among others). These groups frequently cite each other's papers within their fields, but hardly cite other papers in areas beyond their close specialty. A similar situation happens with collaboration networks as well as in other real-world scenarios. Let us just mention a couple of examples from the Stanford Large Network Dataset Collection (SNAP) [36]: the networks representing autonomous system (AS) peering information inferred from "Oregon route-views", "Looking glass data", and "Routing registry", all combined, of March 21th, 2001 (dataset Oregon-2) has $\bar{C} \approx 0.501$ and $C \approx 0.0385$. The second example is the Enron email communication network which has $\bar{C} \approx 0.497$ and $C \approx 0.0302$ (see [2] for details about the previously mentioned networks).

Taking into account the previous considerations we embarked now in a comparison of a real-world network with the windmill graph by considering a series of topological parameters. We consider here both clustering coefficients $\bar{C}$ and C, the maximum degree k_{max}, the size of the largest clique (clique number) Cl, the number of cliques of size Cl in the graph N_{Cl}, the Perron–Frobenius eigenvalue of the adjacency matrix λ_1 and the eigenratio Q. We also study four models frequently used to describe properties of networks. They are the uniform random graphs of Erdős–Rényi (ER) $G_{ER}(n, m)$ [37], the Barabási–Albert (BA) $G_{BA}(n, m_c)$ [35], the Watts–Strogatz small-world model (WS) $G_{WS}(n, m_c)$ [4], and the random geometric graph (RRG) $G_{RRG}(n, r)$ [38]. The main characteristics of these four models are described in [1].

In Table 1 we illustrate the results for the graph parameters and models described before. It is clear that there are several windmill graphs $W(\eta, \kappa)$ which may be used to describe the real-world network under consideration. We selected the windmill graph having the number of edges closer to the one of the real-world network. That is, the number of edges in a windmill graph is

$$m = \frac{\eta\kappa}{2}(\kappa + 1). \tag{34}$$

The graph which better reproduces the number of edges in the real-world network corresponds to the windmill graph $W(33, 7)$, which has 924 edges out of 994 that the citation graph has. As can be seen in Table 1 this graph reproduces very well many of the topological properties of the Small World citation network. It shows a large Watts–Strogatz clustering coefficient and small global transitivity as the real-world network. The highest degree in both the windmill graph and the real-world network are very large in comparison with the number of nodes in these graphs. However, the most interesting results come from the analysis of the clique number and the spectral properties of these graphs. First, the windmill graph is formed by 33 cliques of size 7. That is, the largest clique in this graph is of size 8. The largest clique in the real-world network is of size 7 and there are 39 of them. We should notice that these cliques in the real-world network have overlapping, i.e., there are nodes in more than one clique, which does not occur in the windmill graph. The Perron–Frobenius eigenvalue of the adjacency matrix of the windmill graph is 18.49, which is very close to that of the citation network. Finally, the eigenratio of the windmill, which is just n^{-1}, is identical to that of the real-world network. An inspection of the four other models used for comparison show that none of them captures the topological specificity of the real-world network.

7. Conclusions

The main result of this work is to show that the average WS clustering coefficient and the network transitivity can diverge for certain classes of graphs. The windmill graphs are examples of graphs in which this phenomenon occurs due to the fact that there are many cliques connected to a single node in which no pair of nodes from different cliques are connected. It is important to realize that this clustering divergence also occurs in real-world networks. In particular, it seems to be very frequent in citation and collaboration networks. In citation networks, for instance, there could be a central paper which is highly cited by many other papers. These other papers form clusters, mainly inside a particular discipline, whose authors cite each other. However, the papers in one cluster (discipline) hardly cite a paper in another discipline. Then, it is expected that as these networks grow, the divergence between the clustering coefficients become more marked, resembling more and more the situation taken place in windmill graphs. We have characterized here the spectra of the adjacency and the Laplacian matrices of windmill graphs, which are fundamental for understanding their structure and dynamical properties. We have also shown that many of the structural and dynamical properties of a citation network displaying clustering divergence are reproduced by the appropriate windmill graph. Thus these graphs may represent the core for developing theoretical models of certain kinds of real-life networks.

References

[1] E. Estrada, The Structure of Complex Networks. Theory and Applications, Oxford University Press, 2011.

[2] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167–256.

[3] L.F. Costa, O.N. Oliveira Jr, G. Travieso, F.A. Rodrigues, P.R. Villas Boas, L. Antiqueira, M.P. Viana, L.E. Correa Rocha, Analyzing and modeling real-world phenomena with complex networks: a survey of applications, Adv. Phys. 60 (2011) 329–412.

[4] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.

[5] B. Bollobás, Mathematical results on scale-free random graphs, in: S. Bornholdt, H.G. Schuster (Eds.), Handbook of Graph and Networks: From the Genome to the Internet, Wiley–VCH, Weinheim, 2003, pp. 1–32.

[6] R.D. Luce, A.D. Perry, A method of matrix analysis of group structure, Psychometrika 14 (1949) 95–116.

[7] M.E.J. Newman, The structure of scientific collaboration networks, Proc. Natl. Acad. Sci. USA 98 (2001) 404–409.

[8] S. Wasserman, K. Faust, Social Network Analysis, Cambridge University Press, Cambridge, 1994.

[9] E. Estrada, E. Estrada-Vargas, Distance-sum heterogeneity in graphs and complex networks, Appl. Math. Comput. 218 (2012) 10393–10405.

[10] P. Erdős, A. Rényi, V. Sós, On a problem of graph theory, Studia Sci. Math. Hungar. 1 (1966) 215–235.

[11] J.A. Gallian, A survey: recent results, conjectures, and open problems in labeling graphs, J. Graph Theory 13 (1989) 491–504.

[12] J.F. Wang, F. Belardo, Q.X. Huang, B. Borovicanin, On the largest Q-eigenvalues of graphs, Discrete Math. 310 (2010) 2858–2866.

[13] J.F. Wang, H. Zhao, Q. Huang, Spectral characterization of multicone graphs, Czechoslovak Math. J. 62 (2012) 117–126.

[14] K.C. Das, Proof of conjectures on adjacency eigenvalues of graphs, Discrete Math. 313 (2013) 1925.

[15] A. Abdollahi, S. Janbaz, M.R. Obudi, Graphs cospectral with a friendship graph or its complement, arXiv:1307.5411v1.

[16] J.A. Gallian, Dynamic survey DS6: graph labelling, Electron. J. Combin. DS6 (2007) 1–58.

[17] M. Fiedler, Algebraic connectivity of graphs, Czechoslovak Math. J. 23 (1973) 298–305.

[18] D.M. Cvetković, M. Doob, H. Sachs, Spectra of Graphs: Theory and Application, Academic Press, 1980.

[19] D.M. Cvetković, P. Rowlinson, S. Simic, Eigenspaces of Graphs, Cambridge University Press, 1997.

[20] A. Barrat, M. Barthelemy, A. Vespignani, Dynamical Processes on Complex Networks, Cambridge University Press, 2008.

[21] E. Estrada, Introduction to complex networks: structure and dynamics, in: Evolutionary Equations with Applications in Natural Sciences, Springer International Publishing, 2015, pp. 93–131.

[22] N.T.J. Bailey, The Mathematical Theory of Infectious Diseases and Its Applications, Hafner Press, 1975.

[23] R.M. Anderson, R.M. May, Infectious Diseases in Humans, Oxford University Press, 1992.

[24] J. Marro, R. Dickman, Nonequilibrium Phase Transitions in Lattice Models, Cambridge University Press, 1999.

[25] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, C. Faloutsos, ACM Trans. Inf. Syst. Sec. 10 (2008) 1.

[26] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, Y. Moreno, Europhys. Lett. 89 (2010) 38009.

[27] P. Van Mieghem, J. Omic, R.E. Kooij, IEEE/ACM Trans. Netw. 17 (2009) 1.

[28] M. Boguñá, R. Pastor-Satorras, A. Vespignani, Phys. Rev. Lett. 90 (2003) 028701.

[29] G. Chen, Z. Duan, Network synchronizability analysis: a graph-theoretic approach, Chaos 18 (2008) 037102.

[30] G. Chen, X.F. Wang, X. Li, Fundamentals of Complex Networks: Models, Structures and Dynamics, Wiley, 2015.

[31] A. Trafton, Synchronized brain waves enable rapid learning, MIT News, online: http://newsoffice.mit.edu/2014/synchronized-brain-waves-enable-rapid-learning-0612, June 2014.

[32] Y.-C. Wu, Q. Chaudhari, E. Serpedin, Clock synchronization of sensor networks, IEEE Signal Proc. Mag. 28 (2011) 124–138.

[33] M. Barahona, L.M. Pecora, Synchronization in small-world systems, Phys. Rev. Lett. 89 (2002) 054101.
[34] S. Milgram, The small world problem, Psychology Today 2 (1967) 60–67.
[35] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.
[36] J. Leskovec, Stanford large network dataset collection, 2011.
[37] P. Erdős, A. Rényi, On the evolution of random graphs, Pub. Math. Inst. Hungar. Acad. Sci. 5 (1960) 17–60.
[38] M. Penrose, Random Geometric Graphs, Oxford University Press, 2003.