Dynamics of overlapping structures in modular networks

J. A. Almendral,1 I. Leyva,1 D. Li,2 I. Sendiña–Nadal,1 S. Havlin,2 and S. Boccaletti3,4
1Complex Systems Group, Dept. of Teoría de la Señal y Comunicaciones, Universidad Rey Juan Carlos, Camino del Molino s/n, 28943 Fuenlabrada, Madrid, Spain
2Department of Physics, Minerva Center, Bar Ilan University, Ramat Gan 52900, Israel
3Embassy of Italy in Tel Aviv, 25 Hamered Street, 68125 Tel Aviv, Israel
4CNR–Istituto dei Sistemi Complessi, Via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy

(Received 8 May 2010; published 23 July 2010)

Modularity is a fundamental feature of real networks, being intimately bounded to their functionality, i.e., to their capability of performing parallel tasks in a coordinated way. Although the modular structure of real graphs has been intensively studied, very little is known on the interactions between functional modules of a graph. Here, we present a general method based on synchronization of networking oscillators, that is able to detect overlapping structures in multimodular environments. We furthermore report the full analytical and theoretical description on the relationship between the overlapping dynamics and the underlying network topology. The method is illustrated by means of a series of applications.

DOI: 10.1103/PhysRevE.82.016115 PACS number(s): 89.75.Hc, 05.45.Xt

I. INTRODUCTION

Many real systems can be represented by networks made of a large set of interacting elements [1]. The connectivity structure of those graphs often features an organization in modules (clusters, communities), indicating the existence of specialized groups of nodes that are performing a specific role (social, biological, or technological). That is the reason why much effort has been devoted, so far, to create proper tools, able to unveil the network modularity in a large variety of fields [2,3]. However, most of the existing methods provide a full partition of the graph, i.e., they do not consider the possibility of one node belonging to more than one cluster. As a result, those approaches are not suitable to capture the information on how the set of communities interacts.

On the other hand, recent studies have highlighted the crucial importance of modular overlapping, as a possible mechanism through which networking systems are able to perform parallel tasks in a coordinated way [4]. For instance, in a metabolic network, a given metabolite can be involved in more than one function, or, in social networks, a person can belong to several clubs or groups. Actually, only few studies have addressed the issue of finding overlapping structures in complex graphs, with the aim of overcoming the above limitation, and allowing nodes to simultaneously belong to two or more communities, so as to act as interaction bridges between them.

The first approach was based on clique percolation [4], where a community was defined as the set of nodes that can be visited by rolling a k clique over the network through other cliques with k−1 common nodes. Some other methods [5–7] have further tried to describe a graph as a hierarchical set of modules, searching, at the same time, for overlapping structures between them. For instance, Lancichinetti et al. have proposed an algorithm to detect overlapping and hierarchical structures by means of a fitness function [8], while Ref. [9] concentrated in the evaluation of the percentage each node is belonging to the different modules, thus, introducing the concept of fuzzy modularity.

Together with these methods, having in common the fact of being purely topological (i.e., being the identification of overlapping independent on functional or dynamical information), other recent approaches have instead considered the dynamical evolution of the graph as a method to reveal and measure its modular partition [10], the overlapping structure among modules [11], as well as their hierarchical organization [12]. In particular, Ref. [13] introduced a strategy to unveil overlapping nodes and their dynamics, with the main focus on the case of two-modules networks.

In this paper, we extend the analysis to the general situation of a multimodular structure by giving a full analytical description of the overlapping behavior, as well as of its relationship with the network topology. Finally, we illustrate the main results with reference to some networks constructed from the observation of real-world interacting systems.

II. METHOD OF DYNAMICAL OVERLAPPING

We start by considering a graph G of N coupled Kuramoto oscillators [14] grouped in two moduli. The evolution of the network dynamics, given by the phases \( \phi_i \) of the oscillators, is described by:

\[
\dot{\phi}_i = \omega_i + \sum_{j \neq i} a_{ij} \sin(\phi_j - \phi_i),
\]

where dots denote temporal derivatives, \( k_i \) is the degree of the \( i \)th oscillator, \( d \) is the coupling strength, and \( a_{ij} \) are the elements of the adjacency matrix of \( G \) (i.e., they are either 1 or 0 depending on whether or not a link exists from node \( j \) incident to node \( i \)). The frequencies of the oscillators are assigned accordingly to their membership to one or the other module. Precisely, let \( A \) and \( B \) be the two moduli in which the graph \( G \) is dissected, then \( \omega_i = \omega_A \) if \( i \in A \), and \( \omega_i = \omega_B \) if \( i \in B \), this way establishing two clusters of frequencies. Notice that the above formalism includes all kinds of non weighted (directed or undirected) networks.

With some links connecting the two moduli, the onset of a
FIG. 1. (Color online) Political books network. (a) The overlapping index $C_i$ of each book in the network. In this figure we individuate, for the three possible simulations, the synchronization interfaces $S_a$ between the three identified communities: $M_1$ (in blue diamonds) representing left oriented books, $M_2$ (in magenta circles) neither right nor left oriented books, and $M_3$ (in red squares) right oriented books. Books belonging to the corresponding synchronization interface $([C_i] < \epsilon=0.05)$ fall within the gray strip. Nodes simultaneously belonging to two or more synchronization interfaces (i.e., the overlapping nodes) are depicted in full color and encircled with the shape and color of the community they are overlapping with. Nodes 48 and 76, whose dynamical behavior corresponds to a community different from the initially assigned one, are encircled in black. Parameters used in Eq. (1) are: $N=105$, $\omega_1=0.2$, $\omega_2=0.8$, and $d=0.8$. (b) Graph showing the political books network and its three communities. Nodes 48 and 76, initially assigned to community $M_2$, behave as belonging to $M_1$ and $M_3$, respectively.

synchronization interface $S_{AB}$ can occur, that is composed of nodes displaying an instantaneous frequency oscillating in time around the mean value of the frequencies in the two clusters. To identify those nodes belonging to $S_{AB}$, the overlapping index can be defined as:

$$C_i := \text{sgn} \left( \min_t (\phi_i(t) - \bar{\omega}) \mid \phi_i(t) - \bar{\omega} \right),$$

being $\bar{\omega}$ the mean of the two frequencies assigned to the two moduli, which allows to monitor how close in time the dynamics of a node gets to $\bar{\omega}$. The smaller the overlapping index is, the more the corresponding node belongs to the interface, therefore a threshold $\epsilon$ can be fixed, and the condition $|C_i| < \epsilon$ can be taken to assign the $i^{th}$ node to $S_{AB}$ [13].

However, in practical cases, modular networks are usually made of more than 2 communities. For instance, the protein-protein interaction network is seen as to be equipped with many functions such as transcription, translation and energy production, where each function corresponds to one community. Therefore, an extension of the above analysis is required for $G$ to be partitioned into $n$ moduli $M_1, M_2, \ldots, M_n$.

The strategy for identifying all the possible functional interfaces consists in integrating the system defined by Eq. (1) $n$ times. Precisely, in each simulation, one of the $n$ moduli is considered as the cluster of frequency $\omega_1$ and the rest of the graph as the second cluster of frequency $\omega_2$. This way, the nodes for which $|C_i| < \epsilon$ are assigned to the $S_s$ synchronization interface between module $M_p$ and the rest of the network.

Finally, a node $i$ is assigned to the overlapping interface $S_{pq}$ between the modules $M_p$ and $M_q$ if, being a node from either $M_p$ or $M_q$, it belongs simultaneously to $S_p$ and $S_q$; that is, a overlapping interface is defined by

$$S_{pq} := \{i \in (M_p \cup M_q) \cap (S_p \cap S_q)\}. \quad (3)$$

III. ANALYSIS OF REAL-WORLD NETWORKS

In order to illustrate the kind of information that can be unveiled by the analysis of the dynamics of overlapping interfaces, we here apply the method to three multimodular real networks. The first is the network of 105 books on U.S. politics purchased at the online bookseller amazon.com. In such a representation, the links hold for frequent co-purchasing of books by the same buyer. Data has been compiled from 2003 to October 2008, and the modularity analysis of the network [15] has revealed a clear partition of the graph into 3 different communities, which can be related to books with similar ideological orientations.

In our case, we focus the attention on the meaning of the overlapping structure features. To that purpose, we simulate three times Eq. (1), where in each simulation oscillators belonging to modulus $M_p$ ($p=1, 2, 3$) have frequency $\omega_2$ and oscillators belonging to the other two moduli are set at frequency $\omega_1$.

The results are reported in Fig. 1, in which the values of $C_i$ for each book, as given by Eq. (2), can be seen for the three simulation trials. The Figure allows to identify directly, for each trial $p$, all nodes belonging to the synchronization interfaces $S_q$ as those for which the corresponding values of $C_i$ fall within the gray strip. The nodes that, furthermore, fulfill the overlapping condition given by Eq. (3) are depicted in full color and encircled with the shape and color corresponding to the community they overlap with.

Interestingly, our algorithm also identifies two nodes (48 and 76) whose dynamical behavior does not fit in with the community in which they were both initially allocated. Ac-
The network was dissected in Ref. [17] into four communities (two principal ones and three minor subdivisions) using a previously proposed clustering algorithm based on the calculation of betweenness scores [3]. The application of our algorithm to this case is reported in Fig. 2, and highlights the existence of one dolphin, SN89, who actually is responsible of the cohesion between the two main communities (blue circles and the rest). It is worth mentioning that other works [8] analyzing the overlapping structure of this biological network also identified this dolphin as the crucial one for the graph connectivity structure. In addition, as long as the coupling strength $d$ is increased in Eq. (1) from $d=0.9$ to $d=1.0$, our method is also able to detect other “important” dolphins. Namely, Bumper, and Beak are in one overlapping interface, and Patchback, Topless, and Haecksel in another.

Finally, as an example of a technological network, we report the application of the method to the airport transportation network in the USA, as in 1997 [18]. In this case, each node corresponds to an airport in the USA, while two nodes are linked if there is a flight connection between them.

The resulting graph is formed by 332 airports and 2126 flight connections. The application of the fast algorithm developed by Newman in [19] provides an optimal partition consisting in a structure of five communities as depicted in Fig. 3: the first community $M_1$ (blue diamonds) comprising the airports of the west and central part, and the second largest community $M_2$ (red squares) grouping the airports from the eastern part. The third $M_3$ (black stars) and fourth $M_4$ (green triangles) communities include the airports from southern-east states (Louisiana, Mississippi, Alabama and Florida) and Montana. Finally, airports in Alaska are grouped in the last community $M_5$ (not shown).

The resulting overlapping structure (see Fig. 3) shows that the largest overlapping interface individuates the border between the main communities $M_1$ and $M_2$, while the rest is made of the overlap between $M_2$ and $M_3$. Notice that, since the overlapping index is a dynamical measure that depends on the coupling strength, the number of overlapping nodes increases with $d$. In this case, the first nodes to enter the overlapping structure are located in the southern-east states, and correspond to the overlap between $M_1$ and $M_2$ [Dannelly Field (272) and Pensacola Regional (286)], along with Kansas City (172), corresponding to the overlap between $M_2$ and $M_3$. These results are in qualitative agreement with those extracted from Ref. [20], even though the initial community assignment does not coincide exactly, due to the use of different community detection algorithms.
IV. ANALYTICAL RESULTS

Let us now move to describe analytically the overlapping structures found in the previous Section. From the viewpoint of their dynamics, overlapping nodes are characterized by a frustrated behavior due to the competition between different oscillatory clusters in the network. Then, in order to give an analytical basis to our findings, we start again from the simple case in which there are only three interacting phase oscillators \[ i \in \{1, 2, 3\} \]. Two of them represent those nodes clearly assigned to either clusters, so we can consider that their dynamics are given just by the frequencies fixed in each community: \[ \phi_1 = \omega_1 \] and \[ \phi_2 = \omega_2 \]. The third oscillator, which models the set of overlapping nodes, evolves accordingly to both its intrinsic dynamics (\[ \alpha_3 \]) and the inputs received from the others:

\[
\dot{\phi}_3 = \omega_3 + \frac{d}{2} \left[ \sin(\phi_2 - \phi_3) + \sin(\phi_1 - \phi_3) \right].
\]  

Whereas the solutions for oscillators 1 and 2 are simply \[ \phi_1 = \phi_1^0 + \omega_1 t \] and \[ \phi_2 = \phi_2^0 + \omega_2 t \], with \( \phi_1^0 \) and \( \phi_2^0 \) initial phases, the solution of Eq. (4) is \[ \phi_3 = \phi_1 - \theta \], if oscillator 3 is initially assigned to cluster 1, or \[ \phi_3 = \phi_2 + \theta \] otherwise.

In both cases, \( \theta \) is a function whose asymptotic behavior can be approximated up to first order in \( d/\Omega \) (see Appendix A) by

\[
\theta \approx \frac{d}{2\Omega} \cos(\Omega t + \Phi),
\]

being \( \Omega := \omega_2 - \omega_1 \) and \( \Phi := \phi_1^0 - \phi_2^0 \).

Despite being a first order approximation in \( d/\Omega \), an inspection of the instantaneous frequencies in Fig. 4, extracted from the American airline transportation network, shows that there is a good agreement with the harmonic behavior we have theoretically derived. Then, we can consider that Eq. (5) can be also used to describe the asymptotic behavior of a phase oscillator in a directed or undirected complex network without weights. Namely, we can approximate the \( i^{th} \) oscillator phase by

\[
\dot{\phi}_i =\omega_i - s_i A_i \sin(\Omega t),
\]

with

\[
A_i := \frac{\left(k_i^{out}\right)^\gamma}{k_i}.
\]

In Fig. 5 we check the validity of this result with the experimentally estimated amplitudes \( A_{exp} \) of the instantaneous frequencies obtained for a specific simulation trial of the American airline transportation network. The choice of
the understanding and design of generic modular networks. Between the number of links incident to it and coming from theoretically exploration, it is found that the participation of each monitoring of the synchronization process. With further the- to identify the entire structure of overlapping by a careful graph modules. The generic method here introduced is able nodes, the dynamical behavior enforced in each one of the have the role of coordinating, by means of the overlapping graph.

Finally, it is worth noticing that, in practice and due to the exponent \( \gamma \) being usually close to 1, the quantity \( k_{\text{out}}/k_i \) can be used to properly identify the overlapping nodes in a graph.

In conclusion, we have shown that networks with multi-modal structures feature a complex set of interfaces, that have the role of coordinating, by means of the overlapping nodes, the dynamical behavior enforced in each one of the graph modules. The generic method here introduced is able to identify the entire structure of overlapping by a careful monitoring of the synchronization process. With further theoretical exploration, it is found that the participation of each node to such an overlapping structure depends on the ratio between the number of links incident to it and coming from different communities, and its degree. We expect, therefore, that the presented findings will be able to shed some light on the understanding and design of generic modular networks. Actually, similar analysis has been carried out in other real world graphs, including protein-protein interaction networks, whose results will be presented elsewhere.

ACKNOWLEDGMENTS

Work partly supported by EU (Contract No. 043309 GABA), the Spanish Ministry of S&T under Project No. FIS2009-07072, the URJC-CM under Project No. 2008-CET-3575, and the Community of Madrid under the R&D Program of activities MODELICO-CM/S2009ESP-1691. We also wish to thank the EU project DAPHNet, ONR, the Israel Science Foundation, Hadar Foundation and the Center of Complexity Science for financial support. Finally, we thank Javier M. Buldó for his fruitful comments.

APPENDIX A: PROOF OF EQ. (5)

Let us consider three interacting phase oscillators with equations \( \phi_1 = \omega_1, \phi_2 = \omega_2, \), and

\[
\dot{\phi}_3 = \omega_3 + 2\left[\sin(\phi_2 - \phi_3) + \sin(\phi_1 - \phi_3)\right],
\]

(A1)
such a latter equation describing the overlapping community. Since the solutions for oscillators 1 and 2 are \( \phi_1 = \phi_0^1 + \omega_1 t \) and \( \phi_2 = \phi_0^2 + \omega_2 t \), with \( \phi_0^1 \) and \( \phi_0^2 \) initial conditions, Eq. (A1) can be written as

\[
\dot{\phi}_3 = \omega_3 + d \sin\left(\frac{\phi_1 + \phi_2 - \phi_3}{2}\right) \cos\left(\frac{\phi_2 - \phi_1}{2}\right),
\]

where we have assumed that oscillator 3 was initially assigned to cluster 2. Its solution is \( \phi_3 = \phi_0^3 + \omega_3 t + \theta \), where \( \theta \) verifies the following ordinary differential equation:

\[
\dot{\theta} = -d \sin\left(\frac{\Omega t + \Phi}{2}\right) \cos\left(\frac{\Omega t + \Phi}{2}\right),
\]

(A2)

being \( \Omega := \omega_3 - \omega_1 \) and \( \Phi := \phi_0^2 - \phi_0^1 \). Notice that if the oscillator 3 is initially assigned to cluster 1, Eq. (A2) remains valid but now the solution is \( \phi_3 = \phi_0^3 + \omega_3 t - \theta \).

Therefore, the combined inputs from oscillators 1 and 2 make oscillator 3 depart from its intrinsic dynamics in a quantity \( \theta \) that depends on the ratio \( d/\Omega \). Then, at first order in \( \theta \), Eq. (A2) takes the form

\[
\dot{\theta} = -d \left[ \sin\left(\frac{\Omega t + \Phi}{2}\right) + \theta \cos\left(\frac{\Omega t + \Phi}{2}\right)\right] \cos\left(\frac{\Omega t + \Phi}{2}\right),
\]

whose formal solution is

\[
\theta = c e^{-\int f(t) dt} + \frac{d}{2} \int e^{\int f(t) dt} \sin(\Omega t + \Phi) dt,
\]

with \( c \) an integration constant and

\[
f(t) := \frac{d}{2\Omega} \left[ (\Omega t + \Phi + \sin(\Omega t + \Phi)) \right].
\]

But since \( f(t) \approx \frac{1}{2} dt \) asymptotically, \( \theta \) is then approximated by

\[
\theta \approx -\frac{d}{2} e^{-\frac{1}{2}\int f(t) dt} \int e^{\int f(t) dt} \sin(\Omega t + \Phi) dt,
\]

which, after integration and some algebraic manipulations, takes the form
\[
\theta = \frac{\cos(\Omega t + \Phi + \varphi)}{\sqrt{1 + \frac{2\Omega}{d}^2}} = \frac{d}{2\Omega} \cos(\Omega t + \Phi),
\]

with \(\varphi = \arctan \left( \frac{d}{2\Omega} \right)\).

**APPENDIX B: PROOF OF EQ. (7)**

If we consider that the approximate \(i\)th oscillator phase is given by

\[
\phi_i = \omega_i + \frac{d}{k_i} \sum_{j=1}^{N} a_{ij} \sin(\phi_j - \phi_i),
\]

one obtains that the left-hand side can be written as follows:

\[
\phi_i = \omega_i - s_{i\alpha} d \sin(\Omega t + \delta_i).
\]

The right-hand side of Eq. (B2) has a sum over the neighbors of node \(i\) that can be always expressed as

\[
\sum_{j=1}^{N} a_{ij} \sin(\phi_j - \phi_i) = \sum_{j \in C_i} \sin(\phi_j - \phi_i) + \sum_{j \notin C_i} \sin(\phi_j - \phi_i),
\]

where \(C_i\) is the set of the indices associated to those neighbors of node \(i\) that belong also to its community.

But, since \(\omega_j = \omega_i\) for all \(j \in C_i\),

\[
\sum_{j \in C_i} \sin(\phi_j - \phi_i) \approx P_i \quad \text{with} \quad P_i = \sum_{j \in C_i} \sin(b_j - b_i),
\]

and since \(\omega_j = -s_{\Omega} \Omega\) for all \(j \notin C_i\),

\[
\sum_{j \notin C_i} \sin(\phi_j - \phi_i) \approx \sum_{j \notin C_i} \sin(-s_{\Omega} \Omega t + b_j - b_i) = \sin(-s_{\Omega} \Omega t) Q_i + \cos(-s_{\Omega} \Omega t) R_i,
\]

with

\[
Q_i := \sum_{j \notin C_i} \cos(b_j - b_i) \quad \text{and} \quad R_i := \sum_{j \notin C_i} \sin(b_j - b_i).
\]

Consequently, up to first order in \(\frac{d}{k_i}\), Eq. (B2) can be approximated by

\[
- s_{i\alpha} d \sin(\Omega t + \delta_i) = \frac{d}{k_i} P_i - s_{i\alpha} \Gamma_i \sin(\Omega t + \Delta_i),
\]

being

\[
\Gamma_i := \sqrt{Q_i^2 + R_i^2} \quad \text{and} \quad \Delta_i := -s_i \arctan \left( \frac{R_i}{Q_i} \right).
\]

Although the specific values of \(P_i\), \(Q_i\), and \(R_i\) depend on the values \(b_j\), it is possible to compute them for two particular cases. At this point, it is important to remark that, although mathematically suggested by Eq. (B1), the \(b_j\)'s are not fixed by the initial conditions of the system. They appear in an approximate solution after a transient time, thus the distribution of \(b_i\)'s is, in general, uncorrelated with the initial distribution of phases, and it is unknown.

The first case in which computation can be performed analytically, is when the distribution of \(b_i\)'s is a delta (i.e., \(b_i = b\) for all node \(i\)). For this distribution, \(P_i = R_i = 0\) and \(Q_i = k_i^{\text{out}}\), where \(k_i^{\text{out}}\) is the number of links that node \(i\) has with nodes out of its own community. Then, since \(\Gamma_i = k_i^{\text{out}}\) and \(\Delta_i = 0\), Eq. (B3) is only verified if

\[
a_i = \frac{k_i^{\text{out}}}{k_i} \quad \text{and} \quad \delta_i = 0.
\]

The second case corresponds, instead, to the opposite situation, i.e., when the distribution of \(b_i\)'s is flattened until having a uniform distribution (i.e., \(b_i \sim U(0,2\pi)\) for all node \(i\)). First, we have to notice that, defining \(v_j := b_j - b_i\),

\[
\Gamma_i^2 = \sum_j \left[ \sin^2 v_j + \cos^2 v_j \right] + \sum_{k \neq j} \left[ \sin v_j \sin v_k + \cos v_j \cos v_k \right]
\]

\[
= k_i^{\text{out}} + 2 \sum_{k > j} \cos(v_k - v_j).
\]

But since \(v_k - v_j = b_k - b_j\) follows a uniform distribution \(U(0,2\pi)\) because it is the difference of two random phases taken from a uniform distribution, \(\cos(v_k - v_j)\) has a symmetric distribution in \([-1,1]\) and, consequently, its mean is zero. Therefore, we obtain that \(\Gamma_i = \sqrt{k_i^{\text{out}}}\) on average.

Furthermore, given that

\[
\tan \left[ \arctan \left( \frac{R_i}{Q_i} \right) + \alpha \right] = \frac{R_i + \tan \alpha}{Q_i + \tan \alpha} = \frac{R_i \cos \alpha + Q_i \sin \alpha}{R_i \sin \alpha - Q_i \cos \alpha},
\]

the following equality holds for any value of \(\alpha\),

\[
\arctan \left( \frac{\sum_j \sin v_j}{\sum_j \cos v_j} \right) + \alpha = \arctan \left( \frac{\sum_j \sin(v_j + \alpha)}{\sum_j \cos(v_j + \alpha)} \right).
\]

But since \(v_j + \alpha\) follows a uniform distribution \(U(0,2\pi)\), that equality is only possible if \(\arctan(R_i/Q_i)\) has a distribution in \([-\pi, \pi]\) that is invariant under any \(\alpha\) shift, which implies that it is a uniform distribution with a vanishing mean value. Then, we have that \(\Delta_i = 0\) on average.

On the other hand, for the same reason for which \(\cos(b_k - b_j)\) has a zero mean, also \(P_i\) has a zero mean and, therefore, Eq. (B3) is verified on average if
\[ a_i = \frac{\sqrt{k_{iout}}}{k_i} \quad \text{and} \quad \delta_i = 0. \quad (B5) \]

Taking into account Eqs. (B4) and (B5), the approximate solution for the \( i^{th} \) instantaneous frequency is \( \phi_i = \omega_i - \delta_i A_i \), with

\[ A_i := d \left( \frac{k_{iout}}{k_i} \right)^\gamma, \]

being \( \gamma \in [0.5, 1] \) an exponent that depends on the distribution of \( b_i \)'s (close to 1 when it is homogeneous and close to 0.5 when it is heterogeneous).