

The 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. Boccaletti^{3,4}

¹*Complex Systems Group, Dept. of Teoría de la Señal y Comunicaciones,*

Universidad Rey Juan Carlos, Camino del Molino s/n, 28943 Fuenlabrada, Madrid, Spain

²*Department of Physics, Minerva Center, Bar Ilan University, Ramat Gan 52900, Israel*

³*Embassy of Italy in Tel Aviv, 25 Hamered St., 68125 Tel Aviv, Israel*

⁴*CNR- Istituto dei Sistemi Complessi, Via Madonna del Piano 10, 50019 Sesto Fiorentino (Fi), Italy*

(Dated: December 8, 2012)

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 multi-modular 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.

Keywords: complex networks, modularity, community, overlapping

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, these 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 2-modules networks.

In this paper, we extend the analysis to the general situation of a multi-modular 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. THE 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 ϕ_i of the oscillators, is described by:

$$\dot{\phi}_i = \omega_i + \frac{d}{k_i} \sum_{j=1}^N a_{ij} \sin(\phi_j - \phi_i), \quad (1)$$

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 $\omega_i = \omega_1$ if $i \in A$ and $\omega_i = \omega_2$ 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 *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:

$$C_i := \text{sgn}[\min_t\{\dot{\phi}_i(t) - \bar{\omega}\}] \min_t\{|\dot{\phi}_i(t) - \bar{\omega}|\} \quad (2)$$

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 ϵ 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, \dots, 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 ω_1 and the rest of the graph as the second cluster of frequency ω_2 . This way, the nodes for which $|C_i| < \epsilon$ are assigned to the S_k synchronization interface between module k and the rest of the network.

Finally, a node i is assigned to the overlapping interface S_{kl} between the modules M_k and M_l if, being a node from either M_k or M_l , it belongs simultaneously to S_k and S_l ; that is, a overlapping interface is defined by

$$S_{kl} := \{i \in (M_k \cup M_l) \cap (S_k \cap S_l)\}. \quad (3)$$

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 multi-modular real networks.

The first is the network of 105 books on US politics purchased at the on-line 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_i ($i = 1, 2, 3$) have frequency ω_1 and oscillators belonging to the other two moduli are set at frequency ω_2 .

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 one of the trials, all nodes belonging to the synchronization interfaces S_i 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 encircled with a ring whose color is that of 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. Actually, each of these two nodes behaves, in all trials, as belonging to a different community (M_1 and M_3 respectively), indicating a misclassification in the original network partition.

The second network is an ecological one that was constructed from observations of a community of 62 bottlenose dolphins (*Tursiops* spp.) over a period of 7 years from 1994 to 2001 in New Zealand [16]. Nodes in the network represent dolphins, and ties between nodes represent associations between dolphin pairs occurring more often than expected by chance (see Fig. 2).

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 (node 40), who actually is responsible for the maintenance of the cohesiveness of the two bigger communities (blue circles and the

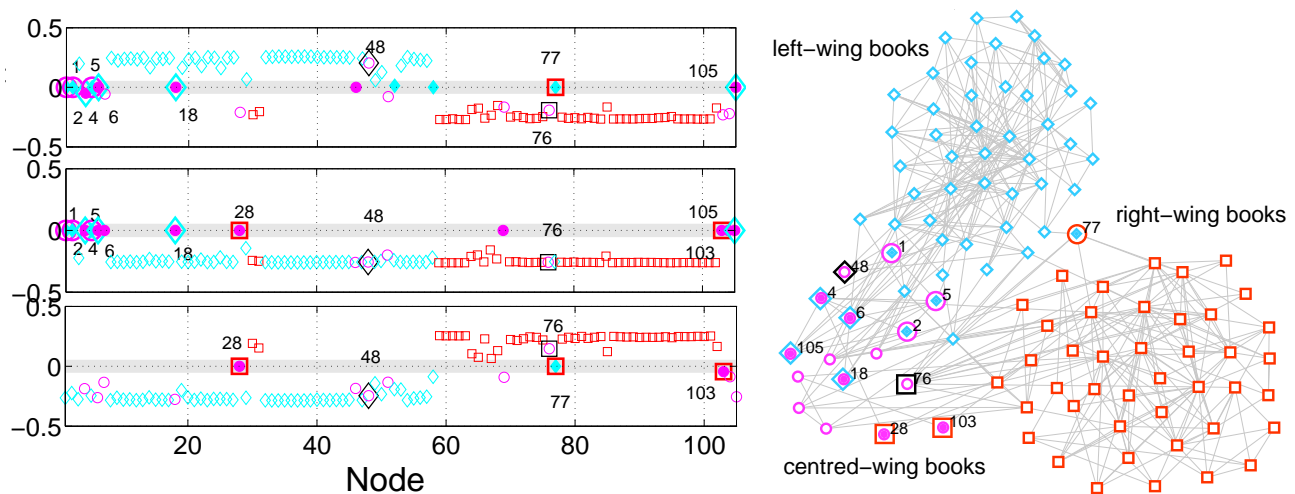


FIG. 1: **Political books network (colors online)**. (A) The overlapping index C_i of each book in the network. In this figure we individuate, for the three possible simulations, the synchronization interfaces S_k between the three identified communities: M_1 (in blue diamonds) representing left oriented 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 encircled with the color of the community they are overlapping with. Nodes 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.

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), our method is also able to detect other “important” dolphins. Namely, DN63 (8), Knit (20), Oscar (29) and PL (31) form overlapping interfaces in the network.

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 2,126 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_3 and M_2 (Dannelly Field (272) and Pensacola Regional (286)), along with Kansas City (172), corresponding to the overlap between M_1 and M_2 .

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 [13]. Two of them represent those nodes clearly assigned to either clusters, so we can consider that their dynamics are given just by the frequencies

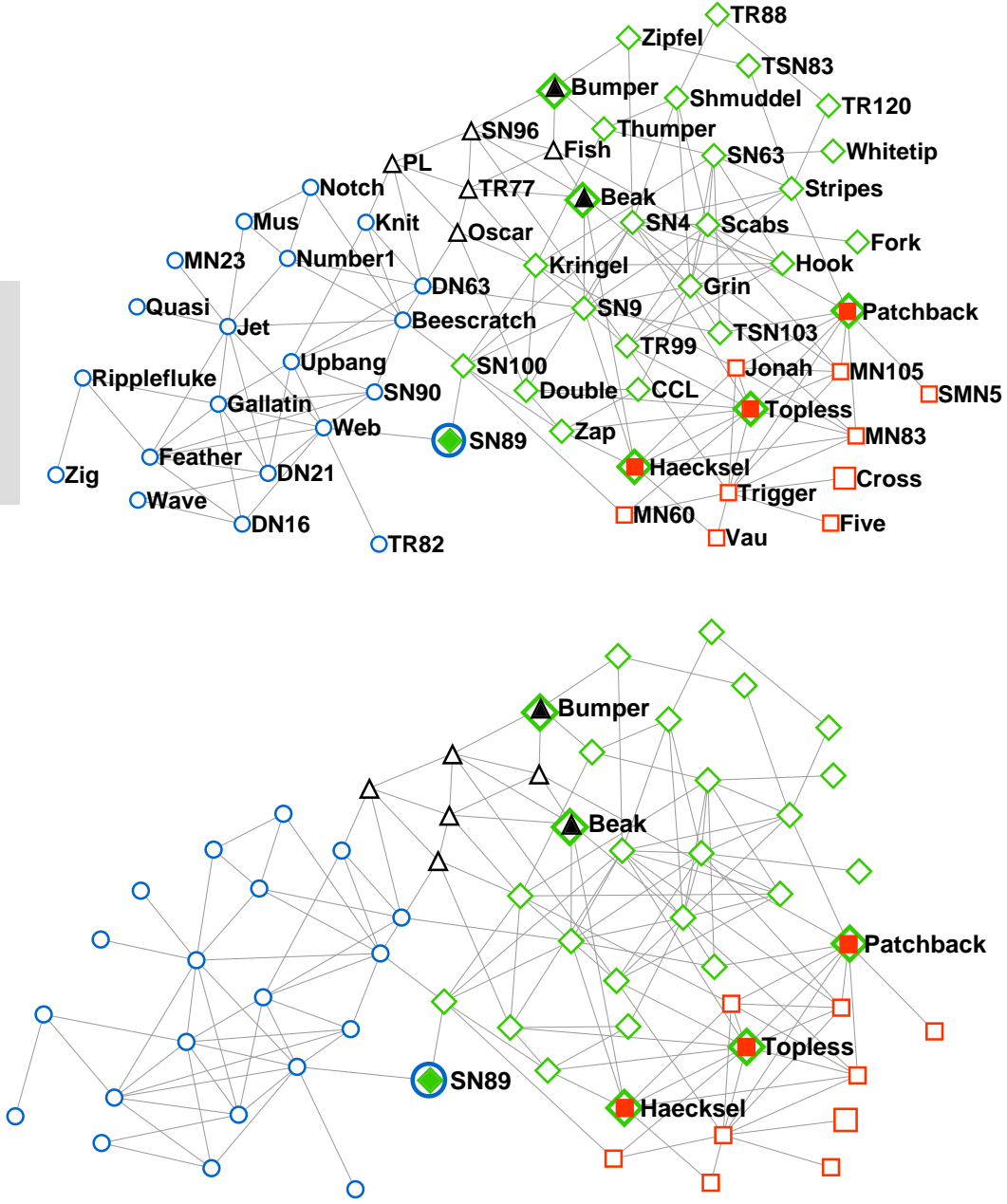


FIG. 2: **Lusseau's bottlenose dolphins network (colors online)**. Graph representation of the four communities given in [17]: there is one big community with 20 individuals (in blue circles), while all other vertices are part of a second big community which is further divided into three sub-communities (with 12, 7 and 23 dolphins) [17]. The identified overlapping individuals with $|C_i| < \epsilon = 0.05$ are full colored and encircled with the symbol corresponding to the community they overlap with. The smallest value of C_i corresponds to a female dolphin named SN89 (green full diamond) who overlaps between the two big communities. Parameters used in Eq. (1) are $N = 62$, $\omega_1 = 0.2$, $\omega_2 = 0.8$, and $d = 1.0$.

fixed in each community: $\dot{\phi}_1 = \omega_1$ and $\dot{\phi}_2 = \omega_2$. The third oscillator, which models the set of overlapping nodes, evolves accordingly to both its intrinsic dynamics (ω_3) and the inputs received from the others:

$$\dot{\phi}_3 = \omega_3 + \frac{d}{2}[\sin(\phi_2 - \phi_3) + \sin(\phi_1 - \phi_3)]. \quad (4)$$

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 ϕ_1^0 and ϕ_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.

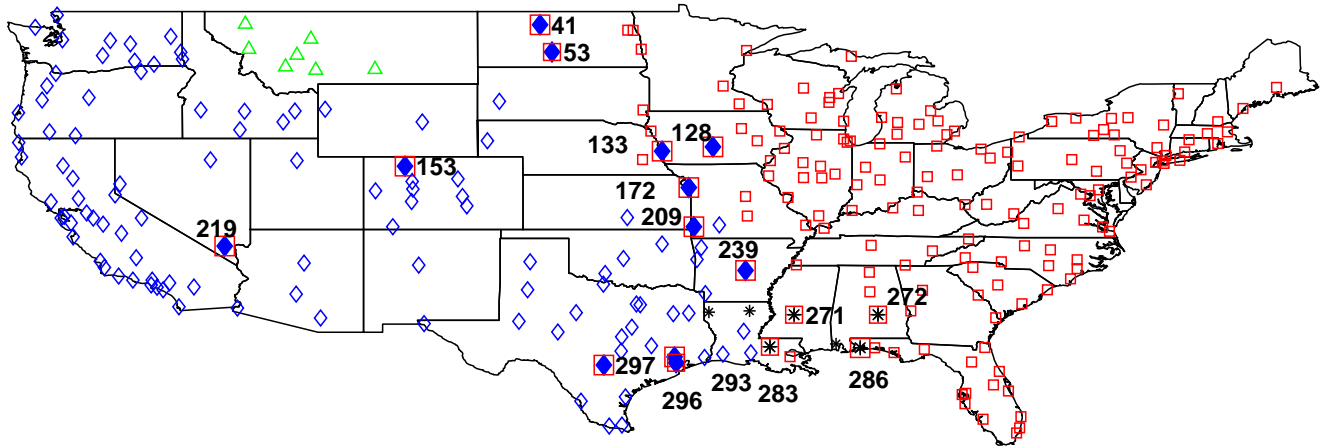


FIG. 3: **American airlines transportation network (colors online)**. Geographical locations of the 5 communities of airports identified in Ref. [19]: M_1 (in blue diamonds) represents the airports located in the western/central states and overseas islands, M_2 (in red squares) the airports located in the eastern states, M_3 (in black stars) those in Louisiana, Mississippi, Alabama and Florida, M_4 (in green triangles) those located in Montana and finally M_5 , not shown in the map, corresponds to airports in Alaska. Those airports whose dynamical overlapping $|C_i| < \epsilon = 0.05$ are encircled with the symbol corresponding to the community they overlap with. Parameters used in Eq. (1) are $N = 332$, $\omega_1 = 0.2$, $\omega_2 = 0.8$, and $d = 0.8$.

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

$$\theta \approx \frac{d}{2\Omega} \cos(\Omega t + \Phi), \quad (5)$$

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

Despite being a first order approximation in $\frac{d}{\Omega}$, an inspection of the instantaneous frequencies in Fig. 4, extracted from the American airlines 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

$$\phi_i = s_i a_i \frac{d}{\Omega} \cos(\Omega t + \delta_i) + b_i + \omega_i t, \quad (6)$$

in which $s_i = -1$ and $\omega_i = \omega_1$ when node i is initially assigned to community 1 ($s_i = +1$ and $\omega_i = \omega_2$ otherwise), and δ_i , a_i , and b_i suitable constants. It is important to remark that, ideally, nodes in the overlapping community are characterized in the former description by $a_i = 0.5$, and the rest by $a_i = 0$. Thus, a_i measures the degree of membership of node i to the overlapping community.

If Eq. (6) is introduced in Eq. (1), we find (see Appendix B) that it is an approximate solution when $\delta_i = 0$ and $a_i = (k_i^{\text{out}})^\gamma / k_i$. k_i is the in-degree of node i . k_i^{out} is the number of links incident to node i from nodes not belonging to the community of node i . On its turn, γ is an exponent in the interval $[0.5, 1]$ that measures how coherent is the asymptotic behavior of the oscillator phases, and depends on the distribution of the set of b_i 's in such a way that it is close to 1 when this phase distribution is homogeneous, and close to 0.5 when it is heterogeneous.

Notice that Eq. (6) rules the i^{th} phase oscillator *after* a transient time, thus the coefficients b_i are not determined by the initial conditions and, consequently, their distribution is unknown in principle.

Therefore, one obtains that the i^{th} instantaneous frequency is

$$\dot{\phi}_i \approx \omega_i - s_i A_i \sin(\Omega t),$$

with

$$A_i := d \frac{(k_i^{\text{out}})^\gamma}{k_i}. \quad (7)$$

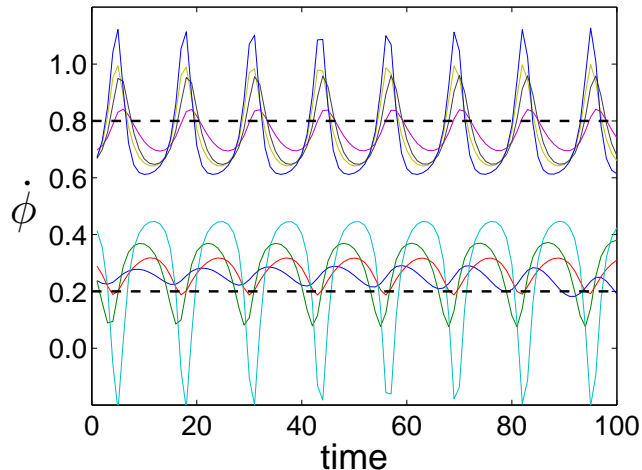


FIG. 4: **Temporal behavior of the instantaneous frequencies of some of the nodes forming the American airlines transportation network (colors online).** Same parameters as in the simulations reported in Fig. 3, where dashed lines are ω_1 and ω_2 . It can be seen how they all qualitatively follow an harmonic oscillation, in agreement with the phases predicted by Eq. 5. The same qualitative behavior is observed for all the other cases considered in the present Manuscript.

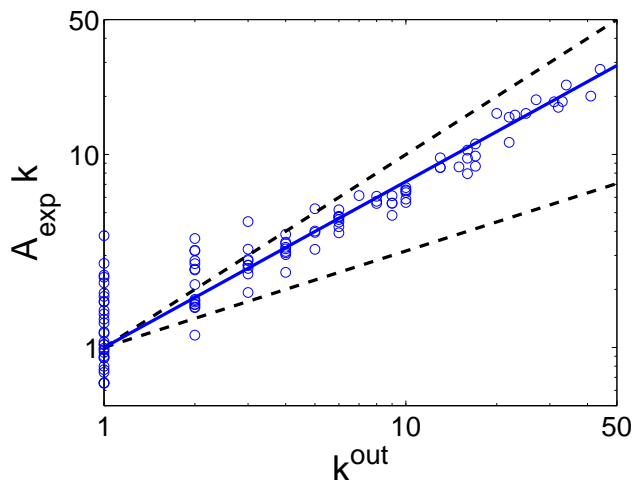


FIG. 5: **Dependence of $A_{exp} k$ with k^{out} in the American airlines transportation network (colors online).** To check Eq. (7), we plot $A_{exp} k$ vs. k^{out} in a log-log scale, being A_{exp} the estimated amplitudes of the instantaneous frequencies in one of the simulation trials. We find that the best fit (blue solid line) is a straight line with slope 0.95, which is between the limiting cases obtained theoretically (black dash lines), corresponding to slopes 0.5 and 1.

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 airlines transportation network. The choice of this particular network is because, due to its high connectivity, the interface between communities is large enough to provide satisfactory statistics.

In that Figure, one can clearly see that the experimental data fit very well Eq. (7) with $\gamma = 0.95$ (blue solid line), which is in between 0.5 and 1 (black dashed lines), as theoretically deduced.

As all trials are considered in Eq. 3, we find that the overlapping nodes (red asterisks in Fig. 6) are mainly concentrated above an horizontal dashed line, corresponding to $A_{exp} = 0.295$. Once this relationship between the overlapping nodes and A_{exp} is established, we further check that the overlapping nodes are those for which $A_i > 0.295$ (on the right of the vertical dashed line). This, on its turn, represents the demonstration that a connection exists between the dynamical evolution of the overlapping nodes and their topological (connectivity) features.

Finally, it is worth noticing that, in practice and due to the exponent γ being usually close to 1, the quantity k_i^{out}/k_i

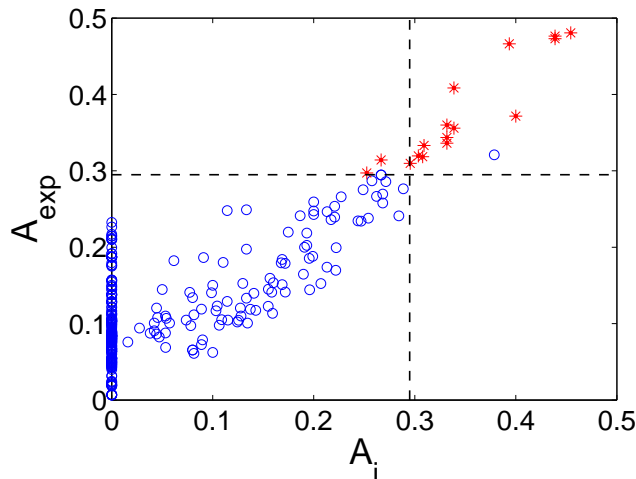


FIG. 6: **Comparison between the theoretical and the numerical amplitudes in the American airlines transportation network (colors online).** The overlapping nodes (red asterisks), characterized by having at least two small overlapping indexes, are essentially those with $A_{exp} \gtrsim 0.3$. But, in addition, they can be rather well described as the nodes with $A_i \gtrsim 0.3$, showing that Eq. 7 determines much of the dynamical behavior exhibited by the nodes in the graph.

can be used to properly identify the overlapping nodes in a graph.

In conclusion, we have shown that networks with multi-modular 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 043309 GABA, the Spanish Ministry of S&T under Project n. FIS2009-07072, the URJC-CM Project 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. Buldú for his fruitful comments.

Appendix A: Proof of Eq. 5

Let us consider three interacting phase oscillators with equations $\dot{\phi}_1 = \omega_1$, $\dot{\phi}_2 = \omega_2$, and

$$\dot{\phi}_3 = \omega_3 + \frac{d}{2}[\sin(\phi_2 - \phi_3) + \sin(\phi_1 - \phi_3)], \quad (\text{A1})$$

such a latter equation describing the overlapping community. Since the solutions for oscillators 1 and 2 are $\phi_1 = \phi_1^0 + \omega_1 t$ and $\phi_2 = \phi_2^0 + \omega_2 t$, with ϕ_1^0 and ϕ_2^0 initial conditions, Eq. A1 can be written as

$$\dot{\phi}_3 = \omega_2 + d \sin\left(\frac{\phi_1 + \phi_2}{2} - \phi_3\right) \cos\left(\frac{\phi_2 - \phi_1}{2}\right),$$

where we have assumed that oscillator 3 was initially assigned to cluster 2. Its solution is $\phi_3 = \phi_2^0 + \omega_2 t + \theta$, where θ verifies the following ordinary differential equation:

$$\dot{\theta} = -d \sin\left(\theta + \frac{\Omega t + \Phi}{2}\right) \cos\left(\frac{\Omega t + \Phi}{2}\right), \quad (\text{A2})$$

being $\Omega := \omega_2 - \omega_1$ and $\Phi := \phi_2^0 - \phi_1^0$. Notice that if the oscillator 3 is initially assigned to cluster 1, Eq. A2 remains valid but now the solution is $\phi_3 = \phi_1^0 + \omega_1 t - \theta$.

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

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

whose formal solution is

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

with c an integration constant and

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

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

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

which, after integration and some algebraic manipulations, takes the form

$$\theta = \frac{\cos(\Omega t + \Phi + \varphi)}{\sqrt{1 + \left(\frac{2\Omega}{d}\right)^2}} \approx \frac{d}{2\Omega} \cos(\Omega t + \Phi),$$

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

$$\phi_i = s_i a_i \frac{d}{\Omega} \cos(\Omega t + \delta_i) + b_i + \omega_i t, \quad (\text{B1})$$

and, if one further introduces the above expression into the Kuramoto model, whose equations are

$$\dot{\phi}_i = \omega_i + \frac{d}{k_i} \sum_{j=1}^N a_{ij} \sin(\phi_j - \phi_i), \quad (\text{B2})$$

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

$$\dot{\phi}_i = \omega_i - s_i a_i d \sin(\Omega t + \delta_i).$$

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

$$\sum_{j=1}^N a_{ij} \sin(\phi_j - \phi_i) = \sum_{j \in \mathcal{C}_i} \sin(\phi_j - \phi_i) + \sum_{j \notin \mathcal{C}_i} \sin(\phi_j - \phi_i),$$

where \mathcal{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 $j \in \mathcal{C}_i$,

$$\sum_{j \in \mathcal{C}_i} \sin(\phi_j - \phi_i) \approx P_i \quad \text{with } P_i := \sum_{j \in \mathcal{C}_i} \sin(b_j - b_i),$$

and since $\omega_j - \omega_i = -s_i \Omega$ for all $j \notin \mathcal{C}_i$,

$$\begin{aligned} \sum_{j \notin \mathcal{C}_i} \sin(\phi_j - \phi_i) &\approx \sum_{j \notin \mathcal{C}_i} \sin(-s_i \Omega t + b_j - b_i) \\ &= \sin(-s_i \Omega t) Q_i + \cos(-s_i \Omega t) R_i, \end{aligned}$$

with

$$Q_i := \sum_{j \notin \mathcal{C}_i} \cos(b_j - b_i) \quad \text{and} \quad R_i := \sum_{j \notin \mathcal{C}_i} \sin(b_j - b_i).$$

Consequently, up to first order in $\frac{d}{\Omega}$, Eq. B2 can be approximated by

$$-s_i a_i d \sin(\Omega t + \delta_i) \approx \frac{d}{k_i} P_i - s_i \frac{d}{k_i} \Gamma_i \sin(\Omega t + \Delta_i), \quad (\text{B3})$$

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_i , 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_i '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 $Q_i = k_i^{out}$, where k_i^{out} is the number of links that node i has with nodes out of its own community. Then, since $\Gamma_i = k_i^{out}$ and $\Delta_i = 0$, Eq. B3 is only verified if

$$a_i = \frac{k_i^{out}}{k_i} \quad \text{and} \quad \delta_i = 0. \quad (\text{B4})$$

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., $b_i \sim \mathcal{U}(0, 2\pi)$ for all node i). First, we have to notice that, defining $\nu_j := b_j - b_i$,

$$\begin{aligned} \Gamma_i^2 &= \sum_j [\sin^2 \nu_j + \cos^2 \nu_j] + \sum_{k \neq j} [\sin \nu_j \sin \nu_k + \\ &\quad + \cos \nu_j \cos \nu_k] = k_i^{out} + 2 \sum_{k > j} \cos(\nu_k - \nu_j). \end{aligned}$$

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

Furthermore, given that

$$\begin{aligned} \tan \left[\arctan\left(\frac{R_i}{Q_i}\right) + \alpha \right] &= \frac{\frac{R_i}{Q_i} + \tan \alpha}{\frac{R_i}{Q_i} \tan \alpha - 1} \\ &= \frac{R_i \cos \alpha + Q_i \sin \alpha}{R_i \sin \alpha - Q_i \cos \alpha}, \end{aligned}$$

the following equality holds for any value of α ,

$$\arctan\left(\frac{\sum_j \sin \nu_j}{\sum_j \cos \nu_j}\right) + \alpha = \arctan\left(\frac{\sum_j \sin(\nu_j + \alpha)}{\sum_j \cos(\nu_j + \alpha)}\right).$$

But since $\nu_j + \alpha$ follows a uniform distribution $\mathcal{U}(0, 2\pi)$, that equality is only possible if $\arctan(R_i/Q_i)$ has a distribution in $[-\pi, \pi]$ that is invariant under any α -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

$$a_i = \frac{\sqrt{k_i^{out}}}{k_i} \quad \text{and} \quad \delta_i = 0. \quad (\text{B5})$$

Taking into account Eqs. B4–B5, the approximate solution for the i^{th} instantaneous frequency is $\dot{\phi}_i \approx \omega_i - s_i A_i \sin(\Omega t)$, with

$$A_i := d \frac{(k_i^{out})^\gamma}{k_i},$$

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).

- [1] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002); M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003); S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Physics Reports* **424**, 175 (2006); S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008); A. Barrat, M. Barthlemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008); R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, 2009).
- [2] M. Girvan and M. E. J. Newman, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821 (2002); F. Radicchi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2658 (2004); R. Guimerá and L. A. N. Amaral, *Nature* **433**, 895 (2005).
- [3] M. E. J. Newman and M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).
- [4] G. Palla, I. Derény, I. Farkas, and T. Vicsek, *Nature* **435**, 814 (2005).
- [5] M. Sales-Pardo, R. Guimerá, A. A. Moreira, and L. A. N. Amaral, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15225 (2007).
- [6] E. N. Sawardeker, M. Sales-Pardo, and L. A. N. Amaral, *Eur. Phys. J. B* **67**, 277 (2009).
- [7] H. Shen, X. Cheng, K. Cai, and M.-B. Hu, *Physica A* **388**, 1706 (2009).
- [8] A. Lancichinetti, S. Fortunato, and J. Kertész, *New J. of Phys.* **11**, 033015 (2009).
- [9] T. Nepusz, A. Petróczi, L. Négyessy, and F. Bacsó, *Phys. Rev. E* **77**, 016107 (2008).
- [10] S. Boccaletti, M. Ivanchenko, V. Latora, A. Pluchino, and A. Rapisarda, *Phys. Rev. E* **75**, 045102(R) (2007).
- [11] T. S. Evans and R. Lambiotte, *Phys. Rev. E* **80**, 016105 (2009).
- [12] X.-H. Wang, L.-C. Jiao, and J.-S. Wu, *Physica A* **388**, 2975 (2009).
- [13] D. Li, I. Leyva, J. A. Almendral, I. Sendiña-Nadal, J. M. Buldú, S. Havlin, and S. Boccaletti, *Phys. Rev. Lett.* **101**, 168701 (2008).
- [14] S. Strogatz, *Physica D* **143**, 1 (2008).
- [15] V. Krebs, *Social Network Analysis software & services for organizations, communities, and their consultants*, <http://www.orgnet.com>.
- [16] D. Lusseau, *Proc. R. Soc. London, Ser. B (Suppl.)* **270**, S186 (2003).
- [17] D. Lusseau and M. E. J. Newman, *Proc. R. Soc. London, Ser. B (Suppl.)* **271**, S477 (2004).
- [18] V. Batagelj and A. Mrvar, *Pajek Datasets*, <http://vlado.fmf.uni-lj.si/pub/networks/data/>.
- [19] M. E. J. Newman, *Phys. Rev. E* **69**, 066133 (2004).
- [20] E. Estrada and N. Hatano, *Appl. Math. and Comp.* **214**, 500 (2009).