IJCSS
International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. **1** (2011), pp. 5–9

Dynamical overlap of protein interaction networks: A method to predict protein functions

I. Sendiña-Nadal[1,2,†], J.A. Almendral[1,2], I. Leyva[1,2], J. M. Buldú[1,2], D. Li[3], S. Havlin[3], Y. Ofran[4] and S. Boccaletti[5,6]

[1] *Complex Systems Group, Rey Juan Carlos University, Camino del Molino s/n, 28943 Madrid, Spain,*

[2] *Laboratory of Biological Networks, CTB - UPM, Campus de Montegancedo 28223 Pozuelo de Alarcón, Madrid, Spain,*

[3] *Dept. of Physics, Minerva Center, Bar Ilan University, 52900 Ramat Gan, Israel,*

[4] *The Mina and Everard Goodman Faculty of Life Sciences, Bar Ilan University, 52900 Ramat Gan, Israel,*

[5] *CNR-Istituto dei Sistemi Complessi, Via Madonna del Piano 10, 50019 Sesto Fiorentino, Firenze, Italy,*

[6] *Embassy of Italy in Tel Aviv, 25 Hamered St., 68125 Tel Aviv, Israel,*

Abstract. While most of the works on functional annotation of proteins via their network of interactions are exclusively based in topological measurements from the properties of the protein interaction network (PIN), we propose the application of an algorithm based on the synchronization behavior emerging from a modular network organization. The method relies on how phase oscillators organize in a network structure of dynamical interactions, and on a recently proposed technique for the identifaction of synchronization interfaces and overlapping communities [5] in ensembles of networking dynamical systems. The combination of the synchronization behavior of the PIN structure and an initial modular classification of proteins allows for protein function predictions of those proteins lying at the overlapping interface that are in agreement with predictions obtained by other methods.

Keywords: complex systems, modularity, overlapping, protein interaction networks

1. Introduction

The latests advances in the field of genoma sequencing technologies have tremendously increase the number of known proteins. The challenge is now how to characterize those proteins and elucidate their function within the different biological processes. One recent approach to assign a function to one protein is by means of the network of its interactions with other proteins [1]. Novel high-throughput techniques for protein-protein interaction measurements have let to obtain those networks of protein interaction from different organisms [2, 3]. Using this network representation, proteins as nodes and detected physical interactions among them as links, it is possible to apply the tools from complex networks theory to predict and annotate a function to a given protein.

2. The method

We start by considering a graph G of N coupled Kuramoto oscillators [4] grouped in two moduli. The evolution of the network dynamics, given by the phase ϕ_i of the oscillators, is described by:

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

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} (SI) 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:

$$C_i := \operatorname{sgn}\left[\min_t\{\dot{\phi}_i(t) - \bar{\omega}\}\right] \min_t\left\{|\dot{\phi}_i(t) - \bar{\omega}|\right\} \tag{2}$$

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} [5].

However, in practical cases, modular networks are usually made of more than 2 communities. For instance, the PIN 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 modules $M_1, M_2, \ldots, M_n$.

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 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 moduli.

Finally, a node is assigned to the overlapping interface between the module M_k and the module M_l (i.e., $i \in S_{kl}$) if, being a node from either M_k or M_l, it belongs simultaneously to S_k and S_l; that is, we denote the overlapping interfaces by

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

3. Results

Our biological bench-test network is the *Saccharomyces cerevisiae* (budding yeast) protein-protein interaction network (pin) reported in [6]. In a much smaller scale, we have tested the algorithm for finding overlapping proteins in a subset of the *Saccharomyces cerevisiae* core protein network downloaded from the DIP database (Database of Interacting Proteins, version ScereCR20050417) [6]. After removal of all self-connecting links, the chosen subset included 269 undirected links between 83 interacting proteins. In this sub-network, Palla *et al.* determined 10 meaningful k-clique communities (associated with either protein complexes or certain functions) by applying the Clique Percolation Method for $k = 4$.

The dynamical behavior of this network by using the classification found by Palla *et al.* and the framework given by Eq. 1, almost identifies as overlapping proteins the same found in [6] as shown in Fig. 1. In this figure the overlapping index of each protein distinguishes between proteins that are tightly associated with a given funcion and those that are more peripheral to it. In this last case, the method is not only able to detect proteins sharing two functions [Cdc42 (1,5), Zds2 (1,7), Ckb2 (3,8), Spt16 (4,10), and Sir1 and Sir4 (7,10)], but also a triple overlap at the protein Cka1, indicating that this

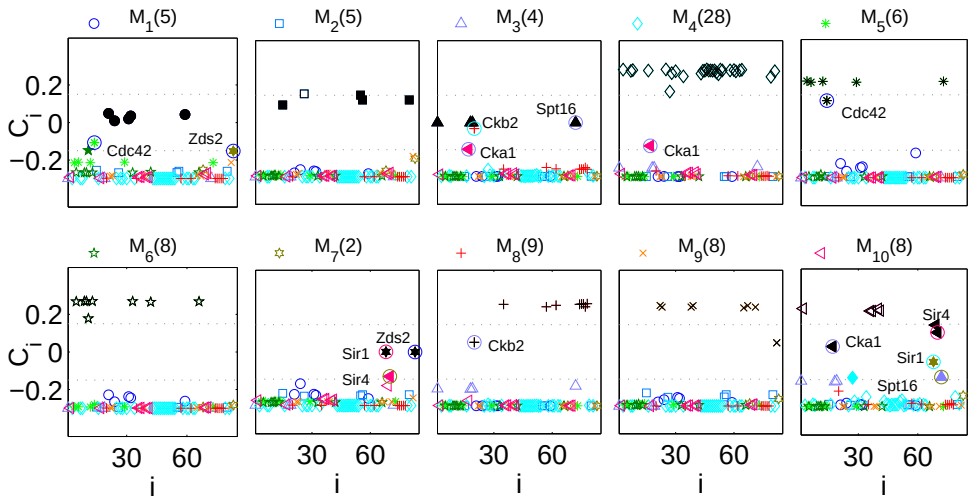

Figure 1: **Identifying the overlapping nodes in a subgraph of the budding yeast pin**. Dynamical overlapping C_i of each node in the network for the 10 possible combinations of clusters of frequencies. Each panel corresponds to the competition between one module M_i (in black symbols) and the rest of modules (in different symbols and colors). The size of each module is written between brackets. Nodes belonging to the corresponding synchronization interface ($|C_i| < 0.15$, blue band) are marked in full face. Those nodes belonging simultaneously to two or more SI, overlapping nodes, are encircled with the color of the corresponding overlapping community. Parameters used in Eq. 1: $N = 83$, $d = 0.65$, $\omega_1 = 0.8$, $\omega_2 = 0.2$.

protein would be involved in the functions denoted by the modules M_3 (transcription regulation), M_4 (ribosome biogenesis), and M_{10} (DNA packaging).

Therefore, the method of dynamical overlapping emerging from the interaction of synchronization interfaces in the modular protein interaction network appears as a valuable tool for accurate function prediction that allows us to assign new function to proteins in that network.

Acknowledgements

Work partly supported by EU contract 043309 GABA, the Spanish Ministry of S&T under Project n. FIS2009-07072, the URJC-CM Project n. 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.

References

[1] R. Sharan, I. Ulitsky, and R. Shamir, *Molecular Systems Biology* **3**, 88 (2007).

[2] R. Aebersold, and M. Mann, *Nature* **422**, 198 (2003).

[3] S. Fields, *FEBS J* **272**, 5391 (2005).

[4] S. Strogatz, *Physica D* **143**, 1-20 (2000).

[5] D. Li et al., *Phys Rev Lett* **101**, 168701 (2008).

[6] G. Palla, I. Derény, I. Farkas, and T. Vicsek, *Nature* **435**, 814 (2005).