



INPE – National Institute for Space Research
São José dos Campos – SP – Brazil – May 16-20, 2016

COMMUNITY DETECTION USING COUPLED KURAMOTO OSCILLATORS WITH CONDITIONAL REPULSION

João Eliakin M. de Oliveira¹, Marcos D. N. Maia², Elbert E. N. Macau³, Marcos G. Quiles⁴

¹INPE - National Institute for Space Research, São José dos Campos, Brazil, joao.eliakin@inpe.br

²INPE - National Institute for Space Research, São José dos Campos, Brazil, mdanielnm@gmail.com

³INPE - National Institute for Space Research, São José dos Campos, Brazil, elbert.macau@inpe.br

⁴UNIFESP - Federal University of São Paulo, São José dos Campos, Brazil, quiles@unifesp.br

Abstract: From Physics to Sociology, complex networks have been presented ubiquitously in the last decade. A particular feature inherited by almost any real-world network is the presence of groups of densely connected nodes, named modules, clusters or communities. In this work, we present a novel model based on coupled Kuramoto oscillators with positive and negative links. Aiming to achieve reliable and efficient community detection, we take advantage of special vertices presented in real networks. This methodology has provided consistent community detection outcomes, which are compared to other related works in literature.

keywords: Complex Networks, Nonlinear Dynamics and Complex Systems, Synchronization in Nonlinear Systems, Community Detection

1. INTRODUCTION

In 1975, the Japanese physicist Yoshiki Kuramoto [1], intrigued by the works of the American biologist Arthur Winfree, on modeling biological phenomena [2], worked on a simple yet complete mathematical model which could express the essential aspects of synchronization, which is shown as follows:

$$\frac{d\theta_i}{dt} = \omega_i + \frac{K}{N} \sum_{j=1}^N a_{ij} \sin(\theta_j - \theta_i), \quad i = 1, \dots, N \quad (1)$$

where θ_i is the phase variable and ω_i is the intrinsic frequency of the i^{th} oscillator (vertex i), N is the total number of oscillators in the system, K is the coupling parameter (coupling strength) which carry the oscillators to a common phase value at the dynamic equilibrium and, finally, a_{ij} is the cell element at the row i and column j of the adjacency matrix $[a_{ij}]$ related to the input complex network.

In its standard form, the Kuramoto model, at the dynamical stability, will not realize communities, since the coupling parameter will make every node in the network evolve to the same phase value (it is called phase locking in the synchronization jargon [3]). In [4], Arenas et al. shown that when applying the Kuramoto model in clustered networks, they will, at first synchronize locally and then, globally, reaching a collective stable state.

In a way to highlight the local synchronization when the system reaches stability, [5], inspired by [6] proposed a modified Kuramoto model as follows:

$$\begin{aligned} \frac{d\theta_i}{dt} = & \omega_i + \frac{K_p}{N} \sum_{j=1}^N a_{ij} \sin(\theta_j - \theta_i) \\ & + \frac{K_n}{N} \sum_{j=1}^N (1 - a_{ij}) \sin(\theta_j - \theta_i) \end{aligned} \quad (2)$$

$i = 1, \dots, N \quad K_p > 0, K_n \leq 0$

in this new form, phases of connected oscillators ($a_{ij} = 1$) are under the original rule, having their phases evolving together but, unconnected oscillators will have the tendency to reach far different phases due to the negative coupling strength.

Since networks with communities are sparse [7], the majority of computation is related to the negative coupling. In Oliveira et al. 2015 [8], an approach for lowering the time complexity of Equation (2) is provided, but lacks on define precisely the group of vertices that will participate as the complement of the network. Thus, the most important contribution of this paper is to provide a better approach for achieving the selection of these vertices.

In summary, this paper is structured as follows: First, in Section 2, its purpose is given, it is shown how the heavy computations are due to the negative coupling and a methodology, from the literature, for lowering the time complexity of Equation (2) is presented. Since this methodology lacks on defining, objectively, the special vertices used for the application of the negative coupling we bring, in Section 3, a new approach that provides results comparable to the ones in the literature, which can be appreciated in Section 4. In Section 5, a conclusion about the outcomes from the experiments done is given, along with possible future works.

2. PURPOSE OF THE PAPER

When storing a graph on a computer, its primitive elements (vertices and edges) have to be placed in memory for fully represent it. By this storage, an adjacency matrix can be generated highlighting which pairs of vertices produces an edge, or not.

Networks with community structure are sparse [7], therefore the number of edges participating in the positive coupling are much less than the number of fictitious edges for the negative one.

Using the *Big O* notation for a graph with n vertices, necessarily applying the positive coupling to the network and the negative coupling to its complement, the dynamics will demand a computational complexity of $O(n^2)$, which is the time complexity of Wu et al.'s approach, presented in Equation (2).

Considering that real networks, e.g. social networks (*Twitter*, *Facebook*), normally are large, with millions of nodes and that their topologies have, most of the time, the presence of community structure and hubs, thus they are indeed sparse, i.e. the number of 1's is less than of 0's in the adjacency matrix.

To make a better use of the computational process, in Oliveira et al. 2015 [8], the authors represented the network with two adjacency lists, one for hubs only (pseudo-adjacency list), and the other as traditional. The dynamics equation shown in Equation (2) can be rewritten as follows:

$$\frac{d\theta_i}{dt} = \omega_i + \frac{K_p}{N} \sum_{j \in L(i)} \sin(\theta_j - \theta_i) + \frac{K_n}{N} \sum_{j \in P(i)} \sin(\theta_j - \theta_i) \quad (3)$$

where $L(i)$ represents the adjacency list of vertex i and, whether i is considered a hub, $P(i)$ represents the hub's pseudo-adjacency list of vertex i , otherwise $P(i)$ is empty and the second summation of Equation (3) is ignored. The parameters K_p and K_n are the same as in Equation (2).

By Equation (3), suppose, for a network with n vertices, that each vertex has degree 1, using adjacency list as the data structure, then for the positive coupling, the time complexity will be $O(n\langle k \rangle)$, where $\langle k \rangle$ is the average degree of the vertices in the network, in this particular case, $\langle k \rangle = 1$, therefore $O(n)$. For the negative coupling, the complexity reached will be $O(n(n - \langle k \rangle))$, which yields $O(n^2)$. Nevertheless, applying the dynamics as in Equation (3) leads to a lower time complexity, since hubs are not very frequent in real networks. According to Oliveira et al. 2015 [8], the time complexity of Equation 3 is $O(n\langle k \rangle) + O(h^2)$, where h is the number of hubs presented in the network.

In the next section, we present the proposed approach for better choosing the vertices that will take part of the pseudo-adjacency list P in Equation 3.

3. METHODS

A potential drawback on Oliveira et al.'s methodology (Equation (3)) is the inexistence of an objective definition of what is a hub. The constraint for a vertex to be considered a hub is having an above average degree. In this context, it is possible that relatively small communities will never be detected, if the vertices separating them to the rest of network, is not considered a hub.

Aiming to improve the quality of the detections, a new approach for finding critical vertices, in real networks, is proposed in this paper. We take advantage of Equation (3), however we apply a new look on what was once considered a hub, composing the pseudo-adjacency list P . The first step is searching for articulation points, i.e. vertices that once removed, provide a disconnected network. By the implementation encountered in the *igraph* library [9], the time complexity to find articulation points is linear, which will not damage the improvements in [8]. It is possible that no vertices have this characteristic at first, nevertheless, removing, iteratively, the most connected vertices, in decrescent order, will eventually provide a disconnected network and, by this means, they will, altogether, compose the set of vertices for which the negative coupling of Equation (3) will be applied.

In the next section, experiments on real networks are presented, the detection outcome is compared to the ones from Wu et al. 2012 [5] and Oliveira et al. 2015 [8].

4. RESULTS

4.1. Zachary's Karate Club Network

Our aim, in this section, is to compare the results acquired using articulation points with those presented in Wu et al. 2012 [5] and Oliveira et al. 2015 [8]. Since we are dealing with real networks, it is interesting having a qualitative approach when analysing the results acquired, once different

methodologies may provide different results, which does not mean that one of them is necessarily wrong.

Figures 1 and 2 shows the following networks: The Zachary's Karate Club [10] and The Protein Interaction Network [11] with the colors representing the modules found by applying Wu et al.'s methodology [5]. The difference between Wu et al.'s and Oliveira et al.'s outcomes is perceptible in the first network. With the old fashioned hubs of Oliveira et al.'s approach, it is impossible to detect the orange community, since each and every vertex in that community is linked to the vertex 1 of the community in red, which happens to be a hub. This leads to a merged red-orange community.

The original history behind the construction of this network [10] says that, in fact, we have two communities lead by vertices 1 and 34. However, by no means we could say that the Wu et al.'s outcome is wrong, they were simply able to realize a small community completely linked to a larger one.

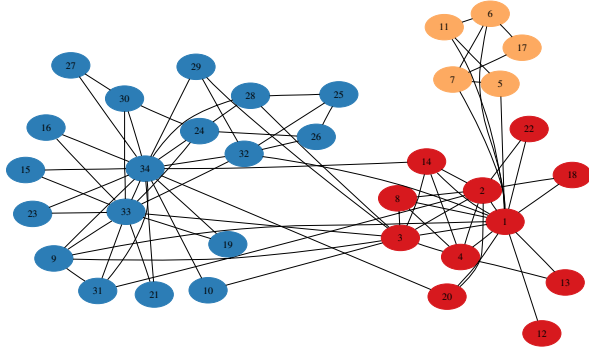


Figure 1 – Zachary's Karate Club Network with its modules detected as in Wu et al. 2012 [5].

With our new approach, the vertex 1 will be identified as an articulation point which makes the orange community separated to the rest of the network. Since the vertex 1 is the only articulation point it is not possible to apply the negative coupling between this vertex and any other. In this context, we start to remove the vertices with the highest number of incident edges, in decrescent order, in this case, vertices 34 and 33. When they are removed altogether, the network is again disconnected. Thus, the vertex 1 will interact, by a negative coupling, with vertices 33 and 34 (these two vertices will also interact negatively, but since they belong to the same community, the influence of other neighbours, by the positive coupling, will compensate).

If the dynamics in Equation (3) is applied the outcome will be the same as in [8], i.e., a merge between communities will happen. However, we are already aware that the vertex 1 is an articulation point, so it is possible to keep the orange community disconnected and apply the dynamics only to the

remaining larger connected component. With this methodology, it is possible to provide two kinds of detections, satisfying both previous methodologies.

The results for the Zachary's Karate Club network can be summarized as follows:

- The vertices of the pseudo-adjacency list P , in Equation (3), will be 1, 33 and 34, applying the negative coupling between them with $K_p = 100.0$ and $K_n = -300.0$ (the parameter values recommended by [8]) we could detect two communities: blue and red-orange as in Oliveira et al. 2015.
- If we consider that 1 is a natural articulation point and putting apart the smaller connected component (the orange community), by applying the dynamics of Equation 3 only on the larger component, we can detect the blue and red communities, as well the orange one.

Therefore, with this new approach, considering articulation points, we may provide outcomes satisfying both previous scenarios.

4.2. Protein Interaction Network

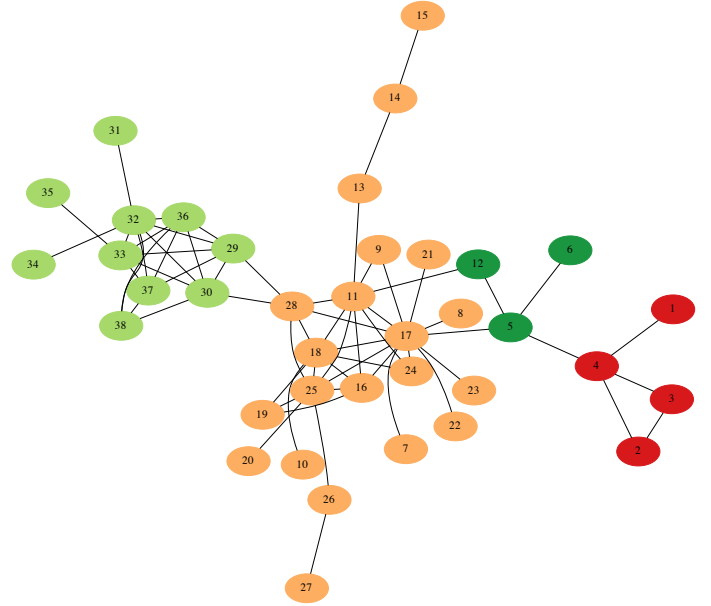


Figure 2 – Protein Interaction network with its modules detected as in Wu et al. 2012 [5].

For the protein interaction network, articulation points are even more remarkable. The set of articulation points are 28, 5 and 4. We do not include articulation points that break the network in only individual vertices or in a branch (set of connected vertices with degree 1, which is the case of vertex 11).

In Figure 3, it is presented the three modules found by Wu et al. 2012, colored as light green, orange and red. We highlighted a group of vertices, in dark green, in which the vertex 5 is a fundamental articulation point to guarantee the detection of the small community in red. Although in Oliveira et al. 2015 the same division in communities were acquired, they have considered 12 hub-vertices and with this new approach, the natural articulation points were sufficient to compose the pseudo-adjacency list P of Equation (3). Therefore, only 3 vertices were necessary to participate on the negative coupling.

The parameters used on the experiments are the same as in Oliveira et al. 2015, i.e. $K_p = 150.0$ and $K_n = -300.0$.

By this two examples, it is clear that articulation points are more decisive than generic hubs, for making more efficient the application of the negative coupling, in Equation (3).

This new approach has almost make irrelevant the factor $O(h^2)$ (see Section 2) in the time complexity of Equation (3), which leads to an even closer linear complexity, since the complexity to find articulation points is indeed linear, as discussed in Section 3.

5. CONCLUSION

In this paper, we aimed to improve the methodology presented by Oliveira et al. 2015 [8] having a different view on which vertices would participate on the negative coupling of the dynamics in Equation (3). Instead of using generic hubs, we took into account articulation points, eventually together with a set of highly connected vertices, which, altogether, provided the same critical effect (disconnecting the original network).

With this new pseudo-adjacency list P , we not only heavily diminished the number of vertices participating on the negative coupling, but we also gave a deeper meaning of what P stands for and, concomitant to it, we were able to provide a more flexible output as for the Karate Club network, where the orange community could be detected or not, according on how the effects of the articulation point is applied in the algorithm.

We believe that the articulation points could be used to cleverly separate the network, *a priori*, into components that are easier to have their communities detected. This feature could be well used in larger networks, then applying the Kuramoto model (Equation (3)) as a refinement.

For future works, an important study would be analyse when should be stopped the removal of most connected vertices, in order to achieve an articulation point effect. In the Karate Club network, it was easy due to its topology. In fact, vertices 1 and 34 are behind its growth, however, for the protein interaction network, we could continue to remove vertices of the orange community. For that network it was sufficient to set as stoppage rule, that would not be considered articulation points, vertices that, once removed, would fragment the network into individual nodes or in a branch, nevertheless it could be possible an otherwise situation.

ACKNOWLEDGMENTS

This paper is supported by the postgraduate program on Applied Computing of the National Institute for Space Research (INPE) with the financial support by the São Paulo Research Foundation (FAPESP), by the Coordination for the Improvement of Higher Education Personnel (CAPES) and by the The National Council for Scientific and Technological Development (CNPq).

References

- [1] Y. Kuramoto and H. Araki, “International symposium on mathematical problems in theoretical physics.,” *Lecture Notes in Physics.*, vol. 39, p. 420, 1975.
- [2] A. T. Winfree, “Biological rhythms and the behavior of populations of coupled oscillators.,” *J. Theor. Biol.*, vol. 16, no. 15, pp. 15–42, 1967.
- [3] A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization : a universal concept in nonlinear sciences*. The Cambridge nonlinear science series, Cambridge: Cambridge University Press, 2001.
- [4] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, “Synchronization in complex networks.,” *Physics Reports*, vol. 469, pp. 93–153, 2008.
- [5] J. Wu, L. Jiao, C. Jin, F. Liu, M. Gong, R. Shang, and W. Chen, “Overlapping community detection via network dynamics.,” *Physical Review.*, vol. E85, no. 016115, pp. 17–61, 2012.
- [6] D. Li, I. Leyva, J. Almendral, I. Sendiña-Nadal, J. Buldú, S. Havlin, and S. Boccaletti, “Synchronization interfaces and overlapping communities in complex networks,” *Phys. Rev. Lett.*, vol. 101, p. 168701, 2008.
- [7] S. Fortunato, “Community detection in graphs,” *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, 2010.
- [8] J. E. M. de Oliveira, M. G. Quiles, E. E. N. Macau, and M. D. N. Maia, “Community detection, with lower time complexity, using coupled kuramoto oscillators.,” *30th Symposium On Applied Computing (SAC) 2015.*, 2015.
- [9] G. Csardi and T. Nepusz, “The igraph software package for complex network research,” *InterJournal*, vol. Complex Systems, p. 1695, 2006.
- [10] W. W. Zachary, “An information flow model for conflict and fission in small groups,” *Journal of Anthropological Research*, pp. 452–473, 1977.
- [11] H.-W. Shen, X.-Q. Cheng, and B.-X. Fang, “Covariance, correlation matrix, and the multiscale community structure of networks,” *Physical Review*, no. 016114, 2010.