

Detecting communities of triangles in complex networks using spectral optimization

Belkacem Serroure^a, Alex Arenas^{b,*}, Sergio Gómez^b

^aLab. LIESP, Bâtiment Nautibus (ex 710), Université Claude Bernard Lyon 1, 69622 Villeurbanne Cedex, France

^bDepartament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain

Abstract

The study of the sub-structure of complex networks is of major importance to relate topology and functionality. Many efforts have been devoted to the analysis of the modular structure of networks using the quality function known as modularity. However, generally speaking, the relation between topological modules and functional groups is still unknown, and depends on the semantic of the links. Sometimes, we know in advance that many connections are transitive and, as a consequence, triangles have a specific meaning. Here we propose the study of the modular structure of networks considering triangles as the building blocks of modules. The method generalizes the standard modularity and uses spectral optimization to find its maximum. We compare the partitions obtained with those resulting from the optimization of the standard modularity in several real networks. The results show that the information reported by the analysis of modules of triangles complements the information of the classical modularity analysis.

Keywords: Complex networks, Communities, Triangle modularity, Spectral optimization

PACS: 89.75.Hc, 02.10.Ox, 02.50.-r

1. Introduction

The study of the modular (or community) structure of complex networks has become a challenging subject [1] with potential applications in many disciplines, ranging from sociology to computer science, see reviews [2, 3, 4]. Understanding the modular units of graphs of interactions (links) between nodes, representing people and their acquaintances, documents and their citation relations, computers and their physical or logical connections, etc., is of utmost

*Corresponding author

Email addresses: bserroure@bat710.univ-lyon1.fr (Belkacem Serroure), alexandre.arenas@urv.cat (Alex Arenas), sergio.gomez@urv.cat (Sergio Gómez)

importance to grasping knowledge about the functionality and performance of such systems. One of the most successful approaches to identify the underlying modular structure of complex networks, has been the introduction of the quality function called *modularity* [5, 6]. Modularity encompasses two goals: (i) it implicitly defines modules as those subgraphs that optimize this quantity, and (ii) it provides a quantitative measure to find them via optimization algorithms. It is based on the intuitive idea that random networks are not expected to exhibit modular structure (communities) beyond fluctuations [7].

A lot of effort has been put into proposing reliable techniques to maximize modularity [8, 9, 10, 11, 12, 13, 14, 15, 16], see review [17]. To a large extent, the success of modularity as a quality function to analyze the modular structure of complex networks relies on its intrinsic simplicity. The researcher interested in this analysis is endowed with a non-parametric function to be optimized: modularity. The result of the analysis will provide a partition of the network into communities such that the number of edges within each community is larger than the number of edges one would expect to find by random chance. As a consequence, each community is a subset of nodes more connected between them than with the rest of the nodes in the network. The user has to be aware of some aspects about resolution limitations that avoid grasping the modular structure of networks at low scales using modularity [18]. The problem can be solved using multiresolution methods [19, 20].

The mathematical formulation of modularity was proposed for unweighed and undirected networks [5] and generalized later to weighted [6] and directed networks [21]. The generalized definition is as follows

$$Q(C) = \frac{1}{2w} \sum_{i=1}^N \sum_{j=1}^N \left(w_{ij} - \frac{w_i^{out} w_j^{in}}{2w} \right) \delta(C_i, C_j) \quad (1)$$

where w_{ij} is the strength of the link between the nodes i and j of the network, $w_i^{out} = \sum_j w_{ij}$ is the strength of links going from i , $w_j^{in} = \sum_i w_{ij}$ is the strength of links coming to j , and the total strength of the network is $2w = \sum_{ij} w_{ij}$. Finally, C_i is the index of the community to which node i belongs to, and $\delta(x, y)$ is the Kronecker function assigning 1 only if $x = y$, and 0 otherwise.

A close look to Eq.(1) reveals that the building block of the community structure we are looking for, within this formulation, is the link between two nodes. Every term in Eq.(1) accounts for the difference, within a module, between the actual existence of a link with weight w_{ij} and the probability of existence of such a link just by chance, preserving the strength distribution.

However, in many cases the minimal and functional structural entity of a graph is not a simple link but a small structure (motif) of several nodes [22]. Motifs are small subgraphs that can be found in a network and that correspond to a specific functional pattern of that network. Statistical over-representation of motifs (compared with the random occurrence of these sub-structures) has been a useful technique to determine minimum building blocks of functionality in complex networks, and several works exploit their identification [22, 23, 24]. Among the possible motifs, the simplest one is the triangle which represents

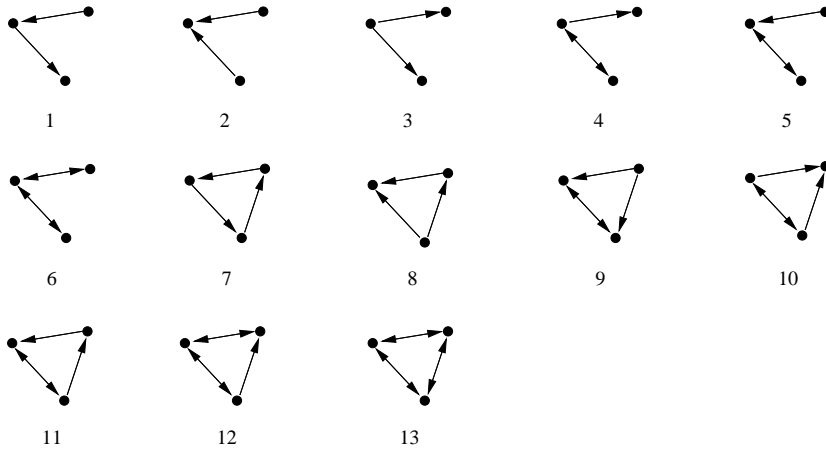


Figure 1: List of all possible three-nodes motifs.

the basic unit of transitivity and redundancy in a graph, see Figure 1. This motif is over-represented in many real networks, for example motifs 12 and 13 in Figure 1, the feedback with two mutual dyads and the fully connected triad respectively, are characteristic motifs of the WWW. Motif 7 (feed-forward loop) is over-represented in electronic circuits, neurons connectivity and gene regulatory transcription networks. The reason for this over-representation relies on the functionality of such small subgraphs on the evolution and performance of the specific network. In the WWW as well as in social networks, the fully connected triad is probably the result of the transitivity of contents or human relations, respectively. The feed-forward loop is related to the reliability or fail tolerance of the connections between important elements involved in communication chains. The idea we propose here is that finding modules containing such motifs as building blocks could improve our information about the modular structure of complex networks. The importance of transitivity is traced back to the seminal paper [25] where it is proposed the clustering coefficient, a scalar measure quantifying the total number of triangles in a network through the average likelihood that two neighbors of a vertex are neighbors themselves.

The main goal of our work is to determine communities using as building blocks triangular motifs. We propose an approach for triangle community detection based on modularity optimization using the spectral algorithm decomposition and optimization. The resulting algorithm is able to identify efficiently the best partition in communities of triangles of any given network, optimizing their correspondent modularity function.

2. Spectral decomposition for triangle community detection

Let $G = (V, A)$ be a weighted undirected graph representing a complex network, where V represents the vertices set and A the edges set. The objective

is to identify communities of triangles, i.e. a partition with the requirement that the density of triangles formed by any three nodes i , j and k inside the same module is larger than the triangles formed outside the module. We will define this objective using a proper adaptation of modularity.

2.1. Triangle modularity tensor

In [26] some of us introduced a mathematical formalism to cope with modularity of motifs of any size. Capitalizing on this work, here we study the specificity of triangle modularity $Q_{\Delta}(C)$ of a certain partition C of an undirected graph (the extension to directed graphs is straightforward, although a little bit more intricate, we present this extension in the Appendix). The mathematical definition is

$$Q_{\Delta}(C) = \sum_i \sum_j \sum_k B_{ijk} \delta(C_i, C_j) \delta(C_j, C_k) \delta(C_k, C_i), \quad (2)$$

where C_i is the index of the community which node i belongs to, and B_{ijk}

$$B_{ijk} = \frac{1}{T_G} w_{ij} w_{jk} w_{ki} - \frac{1}{T_N} (w_i w_j) (w_j w_k) (w_k w_i) \quad (3)$$

is a three indices mathematical object (*triangle modularity tensor*, from now on) that evaluates for each triad i , j , k , the difference between the actual density of strength of the triangle in the graph and the expected density of this triangle in a random configuration with the same strength distribution (null case). The normalization constant T_G is the total number of triads of nodes forming triangles in the network,

$$T_G = \sum_i \sum_j \sum_k w_{ij} w_{jk} w_{ki}, \quad (4)$$

and its counterpart T_N for the null case term is

$$T_N = \sum_i \sum_j \sum_k (w_i w_j) (w_j w_k) (w_k w_i). \quad (5)$$

It is straightforward to check that the triangle modularity tensor satisfies:

$$B_{ijk} = B_{jki} = B_{kij}, \quad (6)$$

$$\sum_i \sum_j \sum_k B_{ijk} = 0. \quad (7)$$

2.2. Spectral optimization of triangle modularity

The computation of the triangle modularity is demanding due to the combinatorial number of triads that can be formed. The proposal of any optimization algorithm for this function must be aware of this cost. Among the possibilities already stated in the literature we devise that the spectral optimization

scheme, first proposed in [16], is a candidate to perform this task efficiently. The idea behind this algorithm is to use the eigenspectrum of the modularity matrix, which plays a role in community detection similar to that played by the graph Laplacian, and use a recursion splitting reminiscent of graph partitioning calculations. The problem we have is that a direct mapping to the usual spectral modularity optimization is not straightforward given the structure of Eq.(2). Basically we need to transform Eq.(2) in a function with the following structure:

$$Q(C) \propto \sum_i \sum_j s_i M_{ij} s_j, \quad (8)$$

where the leading eigenvector of M_{ij} , the modularity matrix, will induce the first recursion step, splitting the network in two parts.

We propose the following transformation: let us assume a partition of the network in two communities, introducing the variables s_i , which are $+1$ or -1 depending on the community to which node i belongs to, and taking into account that

$$\delta(C_i, C_j) = \frac{1}{2}(1 + s_i s_j), \quad (9)$$

then

$$\begin{aligned} \delta(C_i, C_j) \delta(C_j, C_k) \delta(C_k, C_i) &= \frac{1}{8}(1 + s_i s_j)(1 + s_j s_k)(1 + s_k s_i) \\ &= \frac{1}{4}(1 + s_i s_j + s_j s_k + s_k s_i), \end{aligned} \quad (10)$$

where we have made use of $s_i^2 = +1$. Therefore, using Eqs. (6) and (7),

$$\begin{aligned} Q_{\Delta}(S) &= \frac{1}{4} \sum_i \sum_j \sum_k B_{ijk} (1 + s_i s_j + s_j s_k + s_k s_i) \\ &= \frac{3}{4} \sum_i \sum_j \sum_k B_{ijk} s_i s_j. \end{aligned} \quad (11)$$

Defining the *triangle modularity matrix*

$$\begin{aligned} M_{ij} &= \sum_k B_{ijk} \\ &= \frac{1}{T_G} w_{ij} \sum_k w_{jk} w_{ki} - \frac{1}{T_N} (w_i w_i)(w_j w_j) \sum_k (w_k w_k). \end{aligned} \quad (12)$$

then

$$Q_{\Delta}(S) = \frac{3}{4} \sum_i \sum_j s_i M_{ij} s_j. \quad (13)$$

Thus, we have been able to reduce the optimization of the triangle modularity into the standard spectral algorithm given in [16].

For the case of undirected networks, this matrix is symmetric and the computation of its eigenspectra gives real values. However, if the network is directed,

this property is not necessarily true, and then a symmetrization of the matrix is needed before computing its spectrum (see Appendix).

Once a first division of the network in two parts has been obtained, it is possible to iterate the process, while modularity improves, by a recursive application of the spectral splitting to each subgraph. To this end, we need the value of the triangle modularity matrix for any subgraph. Supposing we have a subgraph g to be divided into g_1 and g_2 , the change in triangle modularity is given by

$$\begin{aligned}
\Delta Q_{\Delta}(g \rightarrow g_1, g_2) &= \sum_{i,j,k \in g_1} B_{ijk} + \sum_{i,j,k \in g_2} B_{ijk} - \sum_{i,j,k \in g} B_{ijk} \\
&= \frac{3}{4} \sum_{k \in g} \left(\sum_{i,j \in g} B_{ijk} s_i s_j - \sum_{i,j \in g} B_{ijk} \right) \\
&= \frac{3}{4} \sum_{i,j \in g} s_i M_{ij}(g) s_j, \tag{14}
\end{aligned}$$

where

$$M_{ij}(g) = \sum_{k \in g} \left(B_{ijk} - \delta_{ij} \sum_{\ell \in g} B_{i\ell k} \right), \tag{15}$$

and s_i is $+1$ for nodes in g_1 and -1 for nodes in g_2 . Therefore, the new triangle modularity matrix is not just a submatrix of the original one, but additional terms appear to take into account the connectivity with the rest of the network.

2.3. Algorithm

Once the triangle modularity has been transformed to the proper form to be optimized by spectral decomposition, we can proceed to formulate a complete decomposition-optimization algorithm. After the first analysis of the eigenspectra, the eigenvector associated to the largest eigenvalue is used to determine the elements that will be assigned to one of the two communities according to the sign of their eigenvector component. this process is recursively executed until no new splits are obtained. The decomposition given by the spectral partitioning can be improved by a fine-tuning of the nodes assignments after the process ends.

We use the Kernighan-Lin optimization method to improve the modularity as explained in [16]. The main idea is to move vertices in a group to another increasing the modularity. We move all vertices exactly once. At each step, we choose to move the vertex giving the best improvement (largest increase in the modularity). When all vertices are moved, we repeat the process until no improvement is possible. Some computational issues should be considered here: the computation of the largest eigenvalue and its corresponding eigenvector can be efficiently determined using the iterative Lanczos method [27]; the computation of $Q_{\Delta}(S)$ is, in principle, of order $O(N^3)$, however it can be done very efficiently by pre-computing and storing the values of T_N and T_G , and the

Algorithm 1 Triangle community detection

Require: *Connected network* $G(V,E)$ **Ensure:** *Triangle communities* C , *Triangle modularity of the partition* $Q_{\Delta}(C)$

- 1: Read network
 - 2: Current subgraph $g \leftarrow G$
 - 3: Build modularity matrix $M(g)$
 - 4: Compute $Q_{\Delta}(g)$
 - 5: Compute leading eigenvalue and eigenvector of $M(g)$
 - 6: Decomposition of group g in two groups: g_1 and g_2 , using the signs of eigenvector components
 - 7: Compute the modularity $Q_{\Delta}(g_1, g_2)$ of the initial split of group g
 - 8: Improve $Q_{\Delta}(g_1, g_2)$ using KL optimization between g_1 and g_2
 - 9: Compute the modularity $Q_{\Delta}(g_1, g_2)$ of the split of group g
 - 10: **if** $Q_{\Delta}(g_1, g_2) > Q_{\Delta}(g)$ **then**
 - 11: **goto** 3 with $g \leftarrow g_1$
 - 12: **goto** 3 with $g \leftarrow g_2$
 - 13: **end if**
-

lists of triangles to which each node belongs to; finally, the KL post-processing stage which is eventually the computational bottleneck of the process, must be parameterized according to the number of nodes we pretend to move and the relative improvement of modularity observed.

3. Results

In this section we show the results of the algorithm, applied to several real networks. We have used the following networks:

- Football [1], a network of American football games between Division IA colleges during regular season Fall 2000.
- Zachary [28], a social network of friendships between 34 members of a karate club at a US university in the 1970s.
- Dolphins [29], an undirected social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand.
- Adjnoun [30], adjacency network of common adjectives and nouns in the novel David Copperfield by Charles Dickens.
- Elec s208 [22], benchmark of sequential logic electronic circuit.
- Neurons [31], network of neural connectivity of the nematode *C.elegans*.
- Cortex [32], network of connections between cortical areas in the cat brain.

Network	Nodes	Links	Q	Q_Δ	$\Delta(Q, Q_\Delta)$
Football	115	613	0.604	0.924	0.529
Zachary	34	78	0.419	0.706	0.685
Dolphins	62	159	0.528	0.817	0.547
Adjnoun	112	425	0.308	0.299	-0.029
Elec s208	122	189	0.686	0.998	0.454
Neurons	279	2287	0.405	0.433	0.069
Cortex	55	564	0.372	0.708	0.903

Table 1: Comparison of standard and triangle modularities.

To evaluate the information provided by the new triangle modularity, we perform a comparison with the standard modularity Eq.1. We have developed a comparison in both the values of the optimal modularity, and the partitions obtained.

3.1. Modularities comparison

Table 1 shows the best standard, and triangle modularities found using spectral optimization. We define a new parameter $\Delta(Q, Q_\Delta) = (Q_\Delta - Q)/Q$ that measures the relative difference between both. Positive values of $\Delta(Q, Q_\Delta)$ indicate that the contribution of triangles to communities is larger than standard modularity communities, and the contrary for negative values.

From Table 1 we observe that in Adjnoun, which is almost a bipartite network, the standard modularity is larger than the triangle modularity, in accordance with the absence of these motifs. On the other side, for the Zachary network, a human social network where transitivity is implicit in many acquaintances, the triangle modularity becomes more informative than the standard modularity. Indeed, the optimal standard modularity proposes a decomposition of this network in four groups, while the optimal triangle modularity is achieved for a partition in two groups plus two isolated nodes (nodes 10 and 12) that do not participate in any triangle. Moreover the partition in two groups is in accordance with the observed split of this network after a fight between the administrator and the instructor of the club, see Figure 2.

3.2. Communities comparison

A deeper comparison consist in to analyze the different modules obtained using the standard and triangle modularity. To this end, we need some measures to analyze the difference in the assignments of nodes to modules, taking into account that we will also have different modular partitions. Here, we use two measures, the Normalized Mutual Information (NMI) and the Asymmetric Wallace Index (AW).

In [33] the authors define the NMI to compare two clusterings. The idea is the following: let be a clustering A with c_A communities and a clustering B with c_B communities, and let us define the confusion matrix N whose rows correspond to the communities of the first clustering (A) and columns correspond to the

(a) Triangle modularity (b) Standard modularity

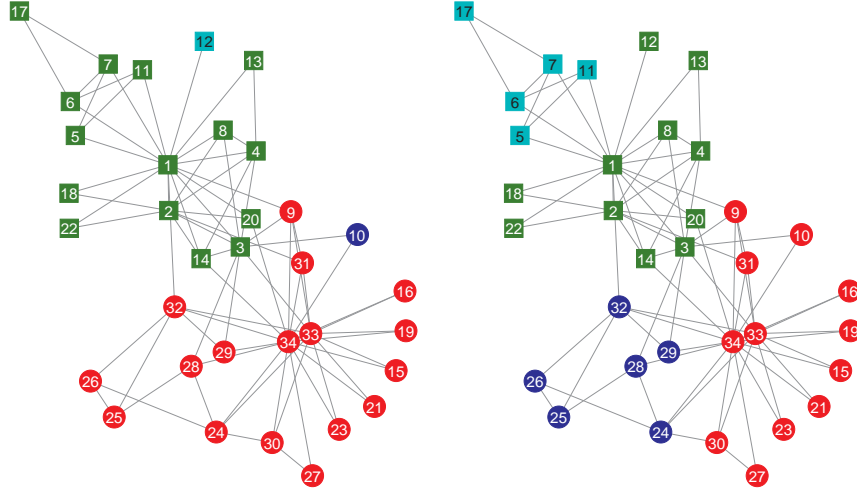


Figure 2: Zachary network partitions. Best partitions found by optimization of (a) triangle modularity and (b) standard modularity. The real splitting of the network is represented by the shape of the symbols (squares and circles). Colors indicate the assignment of nodes to the modules found.

communities of second clustering (B). The elements of the confusion matrix, $N_{\alpha\beta}$, represent the number of common nodes between community α of the clustering A and community β of the clustering B , the partial sums $N_{\alpha.} = \sum_{\beta} N_{\alpha\beta}$ and $N_{.\beta} = \sum_{\alpha} N_{\alpha\beta}$ are the sizes of these communities, and $N_{..} = \sum_{\alpha} \sum_{\beta} N_{\alpha\beta}$ is the total number of nodes. The measure NMI between two clusterings A and B is

$$\text{NMI}(A, B) = \frac{-2 \sum_{\alpha=1}^{c_A} \sum_{\beta=1}^{c_B} N_{\alpha\beta} \log \left(\frac{N_{\alpha\beta} N_{..}}{N_{\alpha.} N_{.\beta}} \right)}{\sum_{\alpha=1}^{c_A} N_{\alpha.} \log \left(\frac{N_{\alpha.}}{N_{..}} \right) + \sum_{\beta=1}^{c_B} N_{.\beta} \log \left(\frac{N_{.\beta}}{N_{..}} \right)} \quad (16)$$

If the partitions are identical, then NMI takes its maximum value of 1. If the partitions are totally independent, $\text{NMI} = 0$. It measures the amount of information that both partitions have in common.

The Asymmetric Wallace Index [34] is the probability that a pair of elements in one cluster of partition A (resp. B) is also in the same cluster of partition B (resp. A). Using the same definitions as for the NMI, the two possible Asym-

Networks	NMI	AW ₁	AW ₂
Football	0.8903	0.8488	0.6901
Zachary	0.6380	0.7945	0.5524
Dolphins	0.6663	0.4810	0.7838
Adjnoun	0.4888	0.3136	0.3845
Elec s208	0.6098	0.0307	0.9091
Neurons	0.6045	0.7276	0.6954
Cortex	0.8361	0.6841	1.0000

Table 2: Comparison of partitions obtained using standard and triangles modularities. The different measures are explained in the text.

metric Wallace Indices are:

$$AW_1(A, B) = \frac{\sum_{\alpha=1}^{c_A} \sum_{\beta=1}^{c_B} N_{\alpha\beta} (N_{\alpha\beta} - 1)}{\sum_{\alpha=1}^{c_A} N_{\alpha} (N_{\alpha} - 1)}, \quad (17)$$

$$AW_2(A, B) = \frac{\sum_{\alpha=1}^{c_A} \sum_{\beta=1}^{c_B} N_{\alpha\beta} (N_{\alpha\beta} - 1)}{\sum_{\beta=1}^{c_B} N_{\beta} (N_{\beta} - 1)}. \quad (18)$$

The asymmetric Wallace index shows the inclusion of a partition in the other.

In Table 2, we observe that the largest NMI is for the communities of football network. That means that the standard and triangle communities found in that network are very similar. Indeed, the structure of the football network is very dense and almost all nodes participate in triangles. For the the AW₂ of the cortex network is equal to 1, that means that all the triangle communities are included in the standard ones.

4. Conclusions

We have designed an algorithm to compute the communities of triangular motifs using an spectral decomposition of the triangle modularity matrix. The algorithm provides partitions where transitive relations are the building blocks of their internal structure. The results of these partitions are complementary to those obtained maximizing the classical modularity, that accounts only for individual links, and can be used to improve our knowledge of the mesoscopic structure of complex networks.

5. Appendix

Here we show the computation of the triangle modularity matrix for a directed motif, in particular motif 7 in Figure 1, although as will be shown the process is equivalent for any other motif configuration. In this case, we have

$$Q_{\Delta}(C) = \sum_i \sum_j \sum_k B_{ijk} \delta(C_i, C_j) \delta(C_j, C_k) \delta(C_k, C_i), \quad (19)$$

where B_{ijk} is

$$B_{ijk} = \frac{1}{T_G} w_{ij} w_{jk} w_{ki} - \frac{1}{T_N} (w_i^{\text{out}} w_j^{\text{in}}) (w_j^{\text{out}} w_k^{\text{in}}) (w_k^{\text{out}} w_i^{\text{in}}). \quad (20)$$

The normalization constant T_G are now

$$T_G = \sum_i \sum_j \sum_k w_{ij} w_{jk} w_{ki}, \quad (21)$$

and

$$T_N = \sum_i \sum_j \sum_k (w_i^{\text{out}} w_j^{\text{in}}) (w_j^{\text{out}} w_k^{\text{in}}) (w_k^{\text{out}} w_i^{\text{in}}). \quad (22)$$

Using the transformation proposed in Eq. (10)

$$\begin{aligned} M_{ij} &= \sum_k B_{ijk} \\ &= \frac{1}{T_G} w_{ij} \sum_k w_{jk} w_{ki} - \frac{1}{T_N} (w_i^{\text{out}} w_i^{\text{in}}) (w_j^{\text{out}} w_j^{\text{in}}) \sum_k (w_k^{\text{out}} w_k^{\text{in}}). \end{aligned} \quad (23)$$

then

$$Q_{\Delta}(S) = \frac{3}{4} \sum_i \sum_j s_i M_{ij} s_j. \quad (24)$$

Owing to the fact that the graph is directed, the modularity matrix M_{ij} may be not symmetric, which causes technical problems. However, it is possible to restore the symmetry thanks to the scalar nature of $Q_{\Delta}(S)$ [35]. A symmetrization of the triangle modularity matrix M ,

$$M' = \frac{1}{2}(M + M^T), \quad (25)$$

yields

$$\begin{aligned} Q_{\Delta}(S) &= \frac{1}{2}(Q_{\Delta}(S) + Q_{\Delta}(S)^T) \\ &= \frac{3}{4} \sum_i \sum_j s_i M'_{ij} s_j, \end{aligned} \quad (26)$$

recovering the necessary symmetry to apply the standard spectral optimization.

In the same manner, we can define the modularity matrix for all possible motifs of Figure 1 just by modifying B_{ijk} . For example, for motif 13 in Figure 1 we have:

$$B_{ijk} = \frac{1}{T_G} w_{ij} w_{ji} w_{jk} w_{kj} w_{ki} w_{ik} - \frac{1}{T_N} (w_i^{\text{out}})^2 (w_j^{\text{in}})^2 (w_j^{\text{out}})^2 (w_k^{\text{in}})^2 (w_k^{\text{out}})^2 (w_i^{\text{in}})^2, \quad (27)$$

$$T_G = \sum_i \sum_j \sum_k w_{ij} w_{ji} w_{jk} w_{kj} w_{ki} w_{ik}, \quad (28)$$

$$T_N = \sum_i \sum_j \sum_k (w_i^{\text{out}})^2 (w_j^{\text{in}})^2 (w_j^{\text{out}})^2 (w_k^{\text{in}})^2 (w_k^{\text{out}})^2 (w_i^{\text{in}})^2. \quad (29)$$

Acknowledgments

We acknowledge J. Borge-Holthoefer and A. Fernández for useful discussions. This work was supported by Spanish Ministry of Science and Technology FIS2009-13730-C02-02 and the Generalitat de Catalunya SGR-00838-2009. B.S. acknowledges support from the Rhone-Alpes region for the financing of training by Explora'doc exchange scholarship.

References

- [1] M. Girvan, M. E. J. Newman, Community structure in social and biological networks, *Proc. Natl. Acad. Sci. USA* 99 (2002) 7821
- [2] L. Danon, A. Díaz-Guilera, J. Duch, A. Arenas, Comparing community structure identification, *J. Stat. Mech.* (2005) P09008
- [3] S. E. Schaeffer, Graph Clustering, *Computer Science Review* 1 (2007) 27-64.
- [4] A. Lancichinetti, S. Fortunato, Community detection algorithms: A comparative analysis, *Phys. Rev. E* 80 (2009) 056117.
- [5] M. E. J. Newman, M. Girvan, Finding and evaluating community structure in networks, *Physical Review E* 69 (2004) 026113.
- [6] M. E. J. Newman, Analysis of weighted networks, *Phys. Rev. E* 70 (2004) 056131.
- [7] R. Guimerà, M. Sales-Pardo, L. A. N. Amaral, Modularity from fluctuations in random graphs and complex networks, *Physical Review E* 70 (2004) 025101.
- [8] M. E. J. Newman, Fast algorithm for detecting community structure in networks, *Physical Review E* 69 (2004) 066133.

- [9] A. Clauset, M. E. J. Newman, C. Moore, Finding community structure in very large networks, *Physical Review E* 70 (2004) 066111.
- [10] L. Donetti and M. A. Muñoz, Detecting network communities: a new systematic and efficient algorithm, *Journal of Statistical Mechanics* 10 (2004) 10012.
- [11] S. Fortunato, V. Latora, M. Marchiori, Method to find community structures based on information centrality, *Physical Review E* 70 (2004) 056104.
- [12] J. Duch, A. Arenas, Community detection in complex networks using extremal optimization, *Physical Review E* 72 (2005) 027104.
- [13] R. Guimerà, L. A. N. Amaral, Functional cartography of complex metabolic networks, *Nature* 433 (2005) 895.
- [14] P. Pons, M. Latapy, Computing communities in large networks using random walks, *Journal of Graph Algorithms and Applications* 10 (2006) 191.
- [15] J. M. Pujol, J. Béjar, J. Delgado, Clustering Algorithm for Determining Community Structure in Large Networks, *Phys. Rev. E* 74 (2006) 016107.
- [16] M. E. J. Newman, Modularity and community structure in networks, *Proceedings of the National Academy of Sciences USA* 103 (2006) 8577.
- [17] S. Fortunato, Community detection in graphs, *Physics Reports* 486 (2010) 75-174.
- [18] S. Fortunato, M. Barthelemy, Resolution limit in community detection, *Proceedings of the National Academy of Sciences USA* 104 (2007) 36.
- [19] J. Reichardt, S. Bornholdt, Detecting fuzzy communities in complex networks with a Potts model, *Phys. Rev. Lett.* 93 (2004) 218701.
- [20] A. Arenas, A. Fernández, S. Gómez, Multiple resolution of the modular structure of complex networks, *New Journal of Physics* 10 (2008) 053039.
- [21] A. Arenas, J. Duch, A. Fernández, S. Gómez, Size reduction of complex networks preserving modularity, *New Journal of Physics* 9 (2007) 176
- [22] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network Motifs: Simple Building Blocks of Complex Networks, *Science* 298 (2002) 824.
- [23] F. Schreiber, H. Schwöbbermeyer, MAVisto: a tool for the exploration of network motifs, *Bioinformatics* 21 (2005) 3572.
- [24] S. Wernicke, F. Rasche., FANMOD: a tool for fast network motif detection, *Bioinformatics* 22 (2006) 1152.
- [25] D. J. Watts, S. H. Strogatz, *Nature* 393 (1998) 440.

- [26] A. Arenas, A. Fernández, S. Fortunato, S. Gómez, Motif-based communities in complex networks, *Journal of Physics A: Mathematical and Theoretical* 41 (2008) 224001.
- [27] B. N. Parlett, H. Simon, L. M. Stringer, On Estimating the Largest Eigenvalue with the Lanczos Algorithm, *Mathematics of Computation* 38 (1982) 15.
- [28] W. W. Zachary, An information flow model for conflict and fission in small groups, *Journal of Anthropological Research* 33 (1977) 452.
- [29] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, S. M. Dawson, The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Can geographic isolation explain this unique trait?, *Behavioral Ecology and Sociobiology* 54 (2003) 396.
- [30] M. E. J. Newman, Finding community structure in networks using the eigenvectors of matrices, *Phys. Rev. E* 74 (2006) 036104.
- [31] <http://www.wormatlas.org/handbook/nshandbook.htm/nswiring.htm>
- [32] J. W. Scannell, C. Blakemore, M. P. Young, Analysis of connectivity in the cat cerebral cortex, *J. Neurosci.* 15 (1995) 1463.
- [33] L. I. Kuncheva, S. T. Hadjitodorov, Using diversity in cluster ensembles, *Proceedings of IEEE Int. Conf. on Systems, Man and Cybernetics*, The Hague, Netherlands, (2004) 1214.
- [34] D. L. Wallace, Comments on A Method for Comparing Two Hierarchical Clusterings, *Journal of the American Statistical Association* 78:383 (1983) 569.
- [35] E. A. Leicht, M. E. J. Newman, Community Structure in Directed Networks, *Physical Review Letters* 100 (2008) 118703.