Extending modularity definition for directed graphs with overlapping communities

V. Nicosia,* G. Mangioni,† V. Carchiolo,‡ and M. Malgeri§
Università di Catania – Dipartimento di Ingegneria Informatica e delle Telecomunicazioni – V.le A. Doria 6 – Catania (ITALY)
(Dated: February 2, 2008)

Complex networks topologies present interesting and surprising properties, such as community structures, which can be exploited to optimize communication, to find new efficient and context–aware routing algorithms or simply to understand the dynamics and meaning of relationships among nodes. Complex networks are gaining more and more importance as a reference model and are a powerful interpretation tool for many different kinds of natural, biological and social networks, where directed relationships and contextual belonging of nodes to many different communities is a matter of fact. This paper starts from the definition of modularity function, given by M. Newman to evaluate the goodness of network community decompositions, and extends it to the more general case of directed graphs with overlapping community structures. Some interesting properties of the proposed extension are discussed and a deeper investigation of metrics and tools for community discovering and interpratation is suggested.

I. INTRODUCTION

The problem of dividing a graph into "interesting" subgraphs is a classical problem in graph theory. Since graphs are used in many application fields to represent different kinds of structures, from relations among people to connections between computers or interconnections among concepts, there are different reasons and motivations for cutting graphs into smaller components. In the field of computer networks, for example, cutting a network (graph) into smaller components is useful to maximise network bandwidth with respect to access to certain services, while in the case of graphs representing communication between processes an optimal subdivision based on minimising the flow among group of processes is of the most importance for optimal scheduling in multi–processors environments.

All those classical problems are often solved by different "clustering methods", which in turn are able to optimise a graph structure in order to guarantee certain desired features, and a huge amount of mathematical literature on that field has been produced.

Nevertheless, studies performed in the last few years on social and natural networks, revealed that algorithms used for graph clustering are neither adapt nor useful to explain partitioning patterns observed in those networks, such as the arising of "communities", "groups" or "clubs". On the other hand, those kind of structures are really interesting, both for theoretical and for practical reasons. First, because they naturally arise as a consequence of simple interactions among people, and do not require complicated mechanisms to be obtained and maintained. Second, because they have some useful properties, such as high internal connectivity, low path length among nodes and high robustness, which are of the most importance in real applications.

A precise definition of what a "community" really is does not exist yet. One of the most widely accepted and used definitions is that given by M. Newman and M. Girvan in [1]: a community is a subgraph containing nodes which are more densely linked to each other than to the rest of the graph or, equivalently, a graph has a community structure if the number of links into any subgraph is higher than the number of links between those subgraphs.

It is not so hard to accept the given definition of communities as a reasonable one: communities in real–life are groups of strongly connected nodes, as happens for example with people in a tennis club, authors in a co–authorship network or colleagues working in the same office [2] [3]. It is worth noting that usually nodes in a community know each other, and the probability for two nodes of a community to have a neighbour in common is higher than for other nodes in the graph [4][5] [6] [7] [8] [9] [10].

Properties of community structures cannot be revealed by classical algorithms for graph clustering: those algorithms are mainly focused on optimal subdivisions of graphs to guarantee min–flow cuts, while finding communities requires a deeper analysis of link patterns and relations. For this reason, a relevant number of new algorithms for community detection have been proposed in the last few years [11][12][13] [14] [15] [16] [17][18] [19] [20] [3] [21] [21] [22].

*Electronic address: vnicosia@diit.unict.it
†Electronic address: gmangioni@diit.unict.it
‡Electronic address: car@diit.unict.it
§Electronic address: mmalgeri@diit.unict.it

Aside with the development of algorithms for community detection, some different metrics for community structure evaluation have been introduced [2] [23], the most popular and widely accepted of those being the so–called "modularity", defined by M. Newman [1] [3].

The modularity seems to be a good measure of community structure, and many algorithms to find graph partitions which give optimal modularity have been proposed so far, and are able to successfully find communities in really large complex networks.

The only drawback of methods based on modularity optimisation is that they gives binary partitions of graphs with respect to vertices. In other words, each vertex can be placed into just one community, and no overlaps among communities are allowed. It is still possible to discover sub–communities, iteratively applying those algorithms to each of the partitions found, but discovering partially overlapped communities is not possible at all.

On the other hand, real complex networks are never divided into sharp sub–networks, especially those formed as a result of social relationships and interactions: people usually belong to many different communities, and participate to activities of a certain number of groups at the same time.

In this paper we propose to extend modularity definition for the more general case of directed graphs with overlapping communities.

The outline of this paper is as follows. In section II we give a brief description of the modularity function as defined by Newman. In section III we discuss our proposal of modularity suitable for discovering overlapped communities in directed graphs. In section IV we test our metric on Zachary karate club network showing that it can be decomposed in two overlapping communities.

II. NEWMAN'S MODULARITY

The idea behind Newman's modularity is simple: a subgraph is a community if the number of links among nodes in the subgraph is higher than expected if links were randomly placed. This is exactly what happens in real–world communities, where the number and density of links among people belonging to groups (families, clubs, user groups etc.) is higher than expected in a random graph of the same size [24][4][5].

This definition of modularity implies the choice of a so–called "null model" [1], i.e. a model of graph to which any other graph can be compared in order to assert the existence of any degree of modularity. When testing for modularity of a complex network, the null model used so far is a random graph with the same number of nodes and edges of the graph to test, but with links among nodes randomly distributed. In such a random graph, the probability P_{ij} of having node i connected to node j is proportional to the degrees (number of links) k_i and k_j of i and j, respectively, and is equal to:

$$P_{ij} = \frac{k_i k_j}{4m^2}, \forall i, j \tag{1}$$

where m is the total number of edges in the graph. If we subdivide an undirected graph $G(E, V)$ into a given number of subgraphs (candidate communities), the modularity of any subgraph $S(E', V') \subseteq G(E, V)$ with respect to the random–graph null model can be computed as the sum of differences between the actual number of links among vertices in V' and the expected number of links among those nodes in the null–model:

$$Q_S = \sum_{i,j \in V'} \left[\frac{A_{ij}}{2m} - P_{ij} \right] \tag{2}$$

where A_{ij} are the terms of the adjacency matrix of $G(E, V)$, defined as:

$$A_{ij} = \begin{cases} 1 & \text{if i and j are connected} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Starting from equation 2, it is possible to define the modularity for the whole graph $G(E, V)$ as follows:

$$Q = \frac{1}{2m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{4}$$

where:

$$\delta(c_i, c_j) = \begin{cases} 1 & \text{if i and j belong to the same community} \\ 0 & \text{otherwise} \end{cases}$$

This formulation of modularity has been successfully used to find or confirm community structures of a relatively large number of networks such as the Zachary karate club network, the dolphins relationship network, collaboration Networks in different research field, and has proved to reasonably catch the structure of subgroups and communities. The problem with this definition is that it is restricted to the particular case of undirected graphs, and it does not take into account possible overlappings among communities in the same network. A reformulation of modularity to deal with directed graphs with overlapping community structure is proposed in the next section.

III. MODULARITY FOR DIRECTED GRAPHS WITH OVERLAPPING COMMUNITIES

In this section we discuss our proposal to extend modularity definition for directed graphs with overlapping communities. We think this generalisation of modularity is needed in order to obtain a metric for evaluation of real–world smooth community structures, as are emerging from sociological, biological and physical studies recently made in the field of complex networks.

A. Modularity for directed graphs

Evaluating the modularity of a given decomposition of a directed graph can be performed in two ways. The first is replacing the adjacency matrix A with a matrix B which has the same dimension of A (i.e. it is an $N \times N$ matrix, where N is the number of nodes) and whose elements $B_{i,j}$ count the number of edges among i and j, so that $B_{i,j} = 0$ if i and j are not connected, while $B_{i,j} = 1$ or $B_{i,j} = 2$ if an edge or two edges (one for each direction) last among i and j, respectively. Note that B is defined as the sum of the adjacency matrix and of its transposed:

$$B = A + A'$$

Replacing A with B, modularity can be computed using the same definition given in equation 4, considering $B_{i,j}$ instead of $A_{i,j}$, but we lack the information about the direction of edges, which is sometimes meaningful to take into account. For example, the case of a node with r out-coming edges and that of a node with $\frac{r}{2}$ incoming nodes and $\frac{r}{2}$ out-coming nodes are completely different from a network topology point of view (in the first case, the given node cannot connect any other couple of nodes, since it does not have any incoming edges), but they are treated in the same way since B discards edge direction.

For this reason, it is possible to reformulate the modularity for directed graphs looking at the meaning of k_i and k_j in equation 4 (similar extensions can be found in [25] and in [26]). The term $\frac{k_i k_j}{4m^2}$ is the probability of having an edge which starts at node i and ends up at node j. This probability is indeed the composition of two different probabilities: the probability that an edge starts at node i, measured as the fraction between the degree of node i and the total number of edges multiplied by two ($\frac{k_i}{2m}$), and the probability that an edge ends at node j, represented, in the same way, $\frac{k_j}{2m}$. Note that for undirected graph there is no difference between the in–degree and the out–degree of a node, and both of them are equal to the node degree, i.e. the number of edges which insist over i.

On the contrary, nodes in a directed graph have out–degree and in–degree which are not the same, since the out–degree is measured as the number of edges that *start* from i and are directed to other nodes, while the in–degree is the number of edges that start from other nodes and *end* up to i. So it is straightforward to redefine the modularity for directed graphs as follows:

$$Q_d = \frac{1}{m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i^{out} k_j^{in}}{m} \right] \delta(c_i, c_j) \tag{5}$$

Note that, in this case, the coefficient $2m$ is replaced by m, since each contribute to modularity is counted just once, while in the original modularity definition it was counted twice since $A_{i,j} = A_{j,i}$ and $k_i^{out} k_j^{in} = k_j^{out} k_i^{in} = k_i k_j$: each undirected edge was implicitly considered as if it was split in two different edges, one for each direction, and hence counted twice (remember that the adjacency matrix on an undirected graph is symmetric). This definition of modularity has the same properties of the original one, since:

- $Q_d = 0$ if all nodes belong to the same community, because $\frac{1}{m}\sum_{i,j\in V}\left[A_{ij} - \frac{k_i^{out}k_j^{in}}{m}\right] = 0$, by construction, and
- Higher values of Q_d indicates a stronger community structure, as happens for the traditional modularity definition, since Q_d has higher values if and only if the distribution of edges inside the communities is higher than that expected in the null–model.

From now on, whenever we refer to modularity we implicitly consider the generalised definition for directed graph given above.

B. Modularity for overlapping graphs

Even if overlapping among communities is an easy–to–understand concept, since overlapping communities can be found in many real networks, an extension of modularity to evaluate the goodness of overlapped community decomposition is a challenging task. Looking at how modularity was first derived by Newman in [1], the first step is to choose a so–called *null–model* to be used as a reference for modularity definition. As reported in section II, Newman states that a network is "modular" when the actual number of connections among nodes in a partition is higher than expected in a corresponding random graph, where the random graph is selected as the reference *null–model*.

In the case of overlapping communities, things are a bit more entangled, because each node can belong to many communities at the same time, and usually it belongs to each community with a certain strength, which is in general not equal for all communities. For this reason, given a directed graph $G(E, V)$ and a set K of overlapped communities built from groups of nodes of G, an array of "belonging factors" $[\alpha_{i,1}, \alpha_{i,2} \ldots \alpha_{i,|K|}]$ can be assigned to each node i in the graph, where each coefficient $\alpha_{i,k}$ expresses how strong node i belongs to community k.

Without loss of generality, we can require that

$$0 \leq \alpha_{i,k} \leq 1 \forall i \in V, \quad \forall k \in K \tag{6}$$

and that

$$\sum_{k=1}^{|K|} \alpha_{i,k} = 1 \tag{7}$$

Note that these positions does affect just the range of belonging factors, not their meaning: a node can belong to many communities at the same time, with different weights. The strength of each belonging is measured as a real value in the range $[0, 1]$ and the sum of all belongings to communities is the same for all nodes in G.

Since each node has a belonging coefficient for each community, it is possible to define a belonging coefficient to each community for edges incoming to or outgoing from a node. We can intuitively suppose that the belonging coefficient to community k of an edge $l = (i, j)$ which starts at node i and ends at node j can be represented by a certain function of the corresponding belonging coefficients of i and j to community k. In formula:

$$\beta_{l,k} = \mathcal{F}(\alpha_{i,k}, \alpha_{j,k})$$

Definition of $\mathcal{F}(\alpha_{i,k}, \alpha_{j,k})$ is somehow arbitrary. It is possible, for example, to define it as the product of the belonging coefficients of involved nodes, or as $\max(\alpha_{i,k}, \alpha_{j,k})$. We actually don't make any choice of a particular form for $\mathcal{F}$.

Even without a precise idea of how a belonging coefficient $\beta_{l,k}$ for a link $l(i, j)$ lasting between i and j can be derived from $\alpha_{i,k}$ and $\alpha_{j,k}$, we can still define the null–model against which a modularity can be estimated.

Note that it is possible to rewrite equation 5 as:

$$Q_d = \frac{1}{m}\sum_{i,j\in V}\left[A_{ij}\delta(c_i, c_j) - \frac{k_i^{out}k_j^{in}}{m}\delta(c_i, c_j)\right] \tag{8}$$

so both the elements A_{ij} of the adjacency matrix and the probability P_{ij} of having a link between i and j in the null model are weighted by the belonging of i and j to the same community, since $\delta(c_i, c_j)$ is equal to 1 only when i and j belong to the same community, and it is 0 otherwise. This formulation allow nodes to belong to only one

community at a time, and the coefficients which multiply A_{ij} and P_{ij} could just be 0 or 1, depending on the fact the i and j really belong to the same community.

Things are a bit different if we consider belonging coefficients $\alpha_{i,k}$ as a measure of how much node i belongs to community k. In this case, each node can belong to many communities at the same time, and its contribution to the modularity of a given community should be weighted by the corresponding belonging coefficient. We can simply reformulate modularity, where $\delta(c_i, c_j)$ is substituted, respectively, by two different coefficients c_{ij} and d_{ij}, obtaining:

$$Q_{ov} = \frac{1}{m} \sum_{i,j \in V} \left[c_{ij} A_{ij} - d_{ij} \frac{k_i^{out} k_j^{in}}{m} \right] \tag{9}$$

It is also possible to put in evidence the contribution to modularity given by each community, so that we can rewrite the modularity as:

$$Q_{ov} = \frac{1}{m} \sum_{k \in K} \sum_{i,j \in V} \left[c_{ijk} A_{ij} - d_{ijk} \frac{k_i^{out} k_j^{in}}{m} \right] \tag{10}$$

We can easily derive a convincing formulation for c_{ijk}. If community belongings are mutually exclusive (the Newman hypothesis), $c_{ijk} = \delta(c_{ik}, c_{jk})$ is the portion of the contribute to modularity given by community k due to link $l(i,j)$ and portion is equal to 1 if and only if i and j are both into the same community k, otherwise it is equal to 0. If we think at c_{ijk} as the weight of the contribution of $l(i,j)$ to modularity of community k, we can define it, in the case of overlapping communities, as the belonging coefficient of $l(i,j)$ to community k:

$$c_{ijk} = \beta_{l(i,j),k} = \beta_{l,k} = \mathcal{F}(\alpha_{i,k}, \alpha_{j,k}) \tag{11}$$

A neat definition of d_{ijk} is a bit more complicated, and requires a clear definition of the *null–model* to be used as reference. We observed that in graphs which have a significant modularity, the modularity of a partition is measured as the difference between the number of links which are "internal" to each community and the number of total links originated by nodes in the partition. This means that for modular graphs the probability that two nodes belong to the same community is higher if those nodes are neighbours. For this reason, a suitable null–model could be a random graph without a community structure, where the probability of a node to belong to any partition is not related to the fact that any another node belongs to the same partition.

Putting it in a clear way, given a graph $G(E,V)$ we choose as null–model a random graph corresponding to $G(E,V)$ where each node has out–degree and in–degree as in the original graph, and where no particular community partition can be derived by structural properties of the graph, i.e. where the probability that a node i belongs to a given community k with a belonging factor $\alpha_{i,k}$ does not depend upon the probability that any other node j in the network does belong to the same community with $\alpha_{j,k}$. The latter condition is equivalent to saying that the expected belonging coefficient of any possible link $l(i,j)$ starting from a node into community k is simply the average of all possible belonging coefficients of l to k, so that:

$$\beta^{out}_{l(i,j),k} = \frac{\sum_{j \in V} \mathcal{F}(\alpha_{i,k}, \alpha_{j,k})}{|V|} \tag{12}$$

Accordingly, the expected belonging coefficient of any link $l(i,j)$ pointing to a node into community k is:

$$\beta^{in}_{l(i,j),k} = \frac{\sum_{i \in V} \mathcal{F}(\alpha_{i,k}, \alpha_{j,k})}{|V|} \tag{13}$$

Those belonging coefficients are used to weight the probability of having, respectively, a link starting at node i and a link pointing to node j. Modularity in the case of overlapped communities can be accordingly formulated as:

$$Q_{ov} = \frac{1}{m} \sum_{k \in K} \sum_{i,j \in V} \left[\beta_{l(i,j),k} A_{ij} - \frac{\beta^{out}_{l(i,j),k} k_i^{out} \beta^{in}_{l(i,j),k} k_j^{in}}{m} \right] \tag{14}$$

Note that the extension of modularity definition to the case of overlapping communities still depends on the choice of $\mathcal{F}(\alpha_{i,k}, \alpha_{j,k})$, i.e. on the way we choose to weight the contribution of each edge to the modularity calculated for community k.

The most important properties required for modularity in the formulation given by Newman were that:

1. $Q = 0$ when all nodes belong to the same community, i.e. when no community structure can be inferred from topological considerations
2. Higher values of Q indicate stronger community structure

Both of these conditions are satisfied by Q_{ov} as well: if all nodes belong to the same community (condition 1), $|K| = 1$ and

$$\forall i \in V \alpha_{i,1} = 1$$

as required in order to satisfy equation 7. At the same time $\mathcal{F}(\alpha_{i,k}, \alpha_{j,k})$, which expresses a belonging coefficient to the unique community k for the edge that links i and j, should give a value of 1 for $l(i,j) \forall i, j \in V$, and this condition is fully satisfied, for instance, when $\mathcal{F}(\alpha_{i,k}, \alpha_{j,k})$ is a simple average among $\alpha_{i,k}$ and $\alpha_{j,k}$, or is the product of the two values or the max between the two. This implies that

$$\beta_{l,k}^{out} = \frac{\sum_{j \in V} \mathcal{F}(\alpha_{i,k}, \alpha_{j,k})}{|V|} = \frac{|V|}{|V|} = 1 \tag{15}$$

and, similarly:

$$\beta_{l,k}^{in} = \frac{\sum_{i \in V} \mathcal{F}(\alpha_{i,k}, \alpha_{j,k})}{|V|} = \frac{|V|}{|V|} = 1 \tag{16}$$

so that $d_{i,j} = 1$ and the modularity simply reduces to:

$$Q_{ov} = \frac{1}{m} \sum_{i,j \in V} \left[A_{i,j} - \frac{k_i^{out} k_j^{in}}{m} \right] = 0 \tag{17}$$

because

$$\sum_{i,j \in V} A_{i,j} = m$$

and

$$\sum_{i,j \in V} \frac{k_i^{out} k_j^{in}}{m} = m$$

So condition 1 is respected. At the same time, condition 2 is satisfied in a network with a modular structure, since the second member of Q_{ov} represents the case of a completely unstructured network, where even the belonging coefficient of a node i to a given community k is absolutely unrelated with the belonging coefficients to the same community of nodes in the neighbourhood of i. This is clearly not the case in real complex network: it is much more probable that a node belonging coefficient to a given community is similar to the belonging coefficients of its neighbours to the same community. So a strong (overlapping) community structure does imply higher values of Q_{ov}, thus satisfying condition 2.

In the next section we propose a somehow seminal empirical exploration of the properties of the reformulated modularity definition, giving also some working examples of Q_{ov} with a particular choice of $\mathcal{F}(\alpha_{i,k}, \alpha_{j,k})$.

IV. AN EXAMPLE: THE ZACHARY KARATE CLUB

One of the most widely studied example of network with a known modularity structure, used in the last years to compare different algorithms for community detection, is the Zachary Karate Club network. The importance of this dataset is due to the fact that a feedback about real community decomposition is available, since the network represents the relationships of people in a karate club, which was split in two different groups when the karate teacher had an argument with the owner of the club and decided to found his own club.

Many recently proposed algorithms for community decomposition are able to effectively place all the 34 nodes of the Zachary network in the "right" community: the Girvan–Newman [1] algorithm, the fast Newman algorithm, and other algorithms based on modularity optimisation put each person in the right community, and the modularity is really higher when all the nodes are divided as they were really divided in reality.

Anyway, different algorithms (especially those based on statistical optimisation and on any kinds of trial–and–error optimisation, such as genetic algorithms or stochastic Markov chains) sometimes fail to put in the right community the nodes labelled as *3* and *10*. These misplacements have been explained by researchers saying that both nodes can be considered as "border–line" nodes from a topological point of view, because the number of links they have to other nodes in the first and in the second community is almost the same, and modularity optimisation is not affected so much by putting either of them in the wrong community.

Our opinion is quite different, indeed. We think that nodes *3* and *10* are somehow the proof that the Zachary Karate club was not so sharply divided into two communities, and that those border–line nodes represent the overlap of the two communities found by means of modularity optimisation.

In order to prove this statement, we used the new definition of modularity given in the previous section to evaluate the modularity of the Zachary network. In particular, we decided to set $\mathcal{F}(\alpha_{i,k}, \alpha_{j,k})$ to a two–dimensional logistic function:

$$\mathcal{F}(\alpha_{i,k}, \alpha_{j,k}) = \frac{1}{(1 + e^{-f(\alpha_{i,k})})(1 + e^{-f(\alpha_{j,k})})} \tag{18}$$

where $f(\alpha_{i,k})$ is a simple linear scaling function:

$$f(x) = 2kx - k \tag{19}$$

Note that $\mathcal{F}(\alpha_{i,k}, \alpha_{j,k})$ is practically zero when both $\alpha_{i,k}$ and $\alpha_{j,k}$ are equal to zero, and it is practically 1 when both $\alpha_{i,k}$ and $\alpha_{j,k}$ are equal to 1, so that condition 1 is respected. The choice of a two–dimensional logistic function for $\mathcal{F}$ is due to the fact that it is a somehow smooth non–linear version of the δ distribution used in the traditional modularity definition, and gives reasonable values for different compositions of $\alpha_{i,k}$ and $\alpha_{j,k}$.

When the scaling function $f(x)$ is set to

$$f(x) = 60x - 30 \tag{20}$$

the value of the modularity function when both nodes *3* and *10* belong to the community 0[27] is 0.746795, while the modularity when both nodes are put in the second community is 0.733015, in accordance with the evidence that node *3* and *10* really belong more strongly to the first community.

Leaving all other belonging coefficients untouched, the modularity $Q_{ov}$for the Zachary network has been measured for all possible couples $(\alpha_{3,0}, \alpha_{10,0})$ obtained when $\alpha_{3,0}$ and $\alpha_{10,0}$ grow from 0.0 to 1.0 with a step of 0.01 (10201 modularity values in total, see figure 1), and the maximum is obtained for $\alpha_{3,0} = 0.81$ and $\alpha_{10,0} = 0.63$ with a value of 0.746796. This value is a bit larger than the value obtained when both nodes are put exclusively into the first community[28], probably meaning that the two communities in the Zachary network are overlapped in correspondence of nodes *3* and *10*. Similar results have been obtained using different coefficients for k in $f(x)$: the modularity Q_{ov} is higher when *3* and *10* are put into the two communities at the same time, with a belonging coefficient to the first community higher than that for the second one for both nodes.

V. CONCLUSIONS

This paper proposed an extension of modularity function for directed graphs with overlapping communities. This generalisation moves from simple considerations about the meaning and structure of the original modularity function, using an enriched null–model which takes into account nodes belongings to more than just one community at the

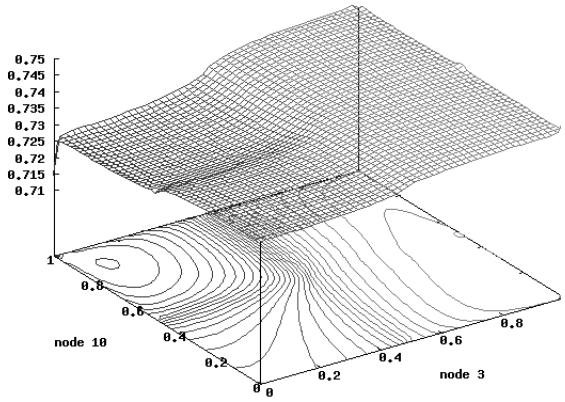

FIG. 1: The generalised modularity function of the Zachary network for different values of $\alpha_{3,0}$ and $\alpha_{10,0}$

same time. The function used to estimate the contribute of an edge to the modularity of a community does not affect the formal procedure used to derive the generalised modularity, and has been left apart as a future research, while a simple application in the case of two dimensional logistic functions has been reported as a proof that the proposed approach is worth to be further studied and investigated.

[1] M. E. J. Newman and M. Girvan, Physical Review E **69**, 026113 (2004), URL http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0308217.

[2] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, Nature **435**, 814 (2005), URL doi:10.1038/nature03607.

[3] M. E. J. Newman, PROC.NATL.ACAD.SCI.USA **103**, 8577 (2006), URL doi:10.1073/pnas.0601602103.

[4] M. E. J. Newman and J. Park, Physical Review E **68**, 036122 (2003), URL http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0305612.

[5] J. Park and M. E. J. Newman, Physical Review E **68**, 026112 (2003), URL http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0303327.

[6] M. Brede and S. Sinha, *Assortative mixing by degree makes a network more unstable* (2005), URL http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0507710.

[7] D. Lusseau and M. E. J. Newman, PROC.R.SOC.LONDON B **271**, S477 (2004), URL http://www.citebase.org/abstract?id=oai:arXiv.org:q-bio/0403029.

[8] P. Holme, Europhysics Letters **64**, 427 (2003), URL http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0308544.

[9] R. Xulvi-Brunet, W. Pietsch, and I. M. Sokolov, Physical Review E **68**, 036119 (2003), URL http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0303252.

[10] M. E. J. Newman and M. Girvan, *Mixing patterns and community structure in networks* (Springer, Berlin, 2003), pp. 66–87, URL http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0210146.

[11] M. Tasgin, A. Herdagdelen, and H. Bingol, *Community detection in complex networks using genetic algorithms* (2007), URL http://www.citebase.org/abstract?id=oai:arXiv.org:0711.0491.

[12] Y. Hu, M. Li, P. Zhang, Y. Fan, and Z. Di, *Community detecting by signaling on complex networks* (2007), URL http://www.citebase.org/abstract?id=oai:arXiv.org:0710.5441.

[13] S. Lehmann, M. Schwartz, and L. K. Hansen, *Bi-clique communities* (2007), URL http://www.citebase.org/abstract?id=oai:arXiv.org:0710.4867.

[14] J. W. Berry, B. Hendrickson, R. A. LaViolette, V. J. Leung, and C. A. Phillips, *Community detection via facility location* (2007), URL http://www.citebase.org/abstract?id=oai:arXiv.org:0710.3800.

[15] V. Gudkov and V. Montealegre, *Community detection in complex networks by dynamical simplex evolution* (2007), URL http://www.citebase.org/abstract?id=oai:arXiv.org:0710.0550.

[16] U. N. Raghavan, R. Albert, and S. Kumara, Physical Review E **76**, 036106 (2007), URL doi:10.1103/PhysRevE.76.036106.

[17] J. Ruan and W. Zhang, *Identifying network communities with a high resolution* (2007), URL http://www.citebase.org/abstract?id=oai:arXiv.org:0704.3759.

[18] K. Wakita and T. Tsurumi, *Finding community structure in mega-scale social networks* (2007), URL http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0702048.

[19] I. Gunes and H. Bingol, *Community detection in complex networks using agents* (2006), URL http://www.citebase.org/abstract?id=oai:arXiv.org:cs/0610129.

[20] M. E. J. Newman, Physical Review E **74**, 036104 (2006), URL doi:10.1103/PhysRevE.74.036104.

[21] A. Clauset, Physical Review E **72**, 026132 (2005), URL http://www.citebase.org/abstract?id=oai:arXiv.org:physics/0503036.

[22] A. Clauset, M. E. J. Newman, and C. Moore, Physical Review E **70**, 066111 (2004), URL http://www.citebase.org/abstract?id=oai:arXiv.org:cond-mat/0408187.
[23] S. Fortunato and M. Barthelemy, *Resolution limit in community detection* PROC.NATL.ACAD.SCI.USA **104**, 36–41(2007)
[24] M. E. J. Newman, SIAM Review **45**, 167 (2003), URL http://www.citebase.org/cgi-bin/citations?id=oai:arXiv.org:cond-mat/0
[25] E. A. Leicht and M. E. J. Newman, *Community structure in directed networks* (2007), URL http://www.citebase.org/abstract?id=oai:arXiv.org:0709.4500.
[26] A. Arenas, J. Duch, A. Fernandez, and S. Gomez, New Journal of Physics **9**, 176 (2007), URL doi:10.1088/1367-2630/9/6/176.
[27] Community 0 is the community which has node 1, i.e. the community which is usually referred as "the first"
[28] All computations have been made in double precision