A new modularity measure for Fuzzy Community detection problems based on overlap and grouping functions

Daniel Gómez[1,*]

Facultad de Estudios Estadísticos. Complutense University, Madrid, Avenida Puerta de Hierro s/n, 28040 Madrid, Spain, E-mail: dagomez@estad.ucm.es

J.Tinguaro Rodríguez, Javier Yáñez*

Facultad de Matemáticas, Complutense University, Madrid, Plaza de las Ciencias 3, 28040 Madrid, Spain. E-mail: jtrodrig@mat.ucm.es, jayage@mat.ucm.es

Javier Montero*

Instituto IGEO-UCM (CSIC), Facultad de Matemáticas, Complutense University, Madrid, Plaza de las Ciencias 3, 28040 Madrid, Spain. E-mail: monty@mat.ucm.es

Abstract

One of the main challenges of fuzzy community detection problems is to be able to measure the quality of a fuzzy partition. In this paper, we present an alternative way of measuring the quality of a fuzzy community detection output based on n-dimensional grouping and overlap functions. Moreover, the proposed modularity measure generalizes the classical Girvan-Newman (GN) modularity for crisp community detection problems and also for crisp overlapping community detection problems. Therefore, it can be used to compare partitions of different nature (i.e. those composed of classical, overlapping and fuzzy communities). Particularly, as usually done with the GN modularity, the proposed measure may be used to identify the optimal number of communities to be obtained by any network clustering algorithm in a given network. We illustrate this usage by adapting in this way a well-known algorithm for fuzzy community detection problems, extending it to also deal with overlapping community detection problems and produce a ranking of the overlapping nodes. Some computational experiments show the feasibility of the proposed approach to modularity measures through n-dimensional overlap and grouping functions.

Keywords: Aggregation operators; overlap functions, grouping functions.

*Corresponding author

Email addresses: dagomez@estad.ucm.es (Daniel Gómez), jtrodriguez@mat.ucm.es, jayage@ucm.es (J.Tinguaro Rodríguez, Javier Yáñez), monty@mat.ucm.es (Javier Montero)

[1]Facultad de Estudios Estadísticos. Complutense University, Madrid, Avenida Puerta de Hierro s/n 28040, Madrid, Spain
email: dagomez@estad.ucm.es

1. Introduction.

Large and complex networks representing relationships among a set of entities have been one of the focuses of interest of scientists in many fields in the recent years. Examples of complex networks include social networks, the world-wide web network, telecommunication networks and biological networks. One of the most important problems in social network analysis is to describe/explain its community structure. Generally, a community in a network is a subgraph whose nodes are densely connected within itself but sparsely connected with the rest of the network.

Community detection problems has been widely studied during the last decade (see e.g. [15, 20]), with many applications to several disciplines. Discovering inherent communities and structures in a social network must be a main objective when we pursue a better understanding of a given network. Nevertheless, real communities in complex networks often present overlap, such that each vertex may occur in more than one community. Community detection problems with overlapping communities have been also studied in the literature (see [38]), with different purposes. On one hand, a main aim of this problem is to uncover communities allowing some key nodes to belong to more than one community. On the other hand, a related aim is to detect and identify those nodes (usually mentioned as *overlapping nodes*) that belong to more than one community. Overlapping nodes may play a special role in a complex network system, and how to detect them is indeed a very interesting issue. In this sense, it is important to remark that most known algorithms, such as divisive algorithms [16] or agglomerative algorithms [15], cannot detect them.

As it is pointed in [23], two distinct types of overlapping are possible: crisp (where each node fully belongs to each community of which it is a member) and fuzzy (where each node belongs to each community up to a different extent or degree). Thus, taking into account this distinction, three classes of community detection problems are possible: classical community detection problems (in which just non-overlapping communities are allowed), crisp overlapping community detection problems (in which a node could belong to more than one community) and fuzzy community detection problems (in which each node has a degree of membership to each community). As a result, there are two main challenges in fuzzy community detection problems. The first is the development of algorithms that produce a fuzzy clustering of the nodes in the network. And the other is to quantify the quality of such a fuzzy partition.

In this paper, we present an alternative way of measuring the quality of a fuzzy community detection output based on n-dimensional grouping and overlap functions [4, 18], that generalize the classical modularity for crisp community detection problems and also for crisp overlapping community detection problems. In addition with this, in this paper we also develop a fuzzy community detection algorithm, an overlapping community detection algorithm and an overlapping-node ranking method. These three proposals will then allow uncovering the fuzzy structure of a network and its overlapping communities (in a crisp way), as well as a procedure to rank the nodes based on this fuzzy structure.

This paper is organized as follows: Section 2 is devoted to recall the basic notions of overlap and grouping functions, both in their bivariate and n-dimensional formulation, as well as to remind the concept of community detection problems, with and without overlapping communities. Similarly, Section 3 reviews the state-of-the-art in modularity measures, allowing the introduction of our new modularity measure for fuzzy community detection problems in Section 4. Our proposed methods for fuzzy community detection and crisp overlapping community detection, as well as the associated ranking process based on overlap and grouping functions are presented in Section 5. Finally, Section 6 is devoted to show the results of some computational experiments and to discuss some concluding remarks.

2. Preliminaries

In this section, we recall some concepts and properties of bivariate and n-dimensional overlap and grouping functions, which were initially proposed in [4, 24], and extended to the n-dimensional case in [18].

2.1. Bivariate overlap and grouping functions

Aggregation is a basic and necessary tool for most knowledge-based systems. An *aggregation operator* [3, 8–12, 21] is usually defined as a real function A_n that, from n data items $x_1, \ldots, x_n$ in $[0,1]$, produces an aggregated value $A_n(x_1, \ldots, x_n)$ in [0,1] [7, 12]. Some desirable properties any aggregation operator should satisfy use to be imposed: for example, some boundary conditions (for all n, $A_n(0, \ldots, 0) = 0$ and $A_n(1, \ldots, 1) = 1$), monotonicity and continuity in each variable (see again [5, 10]. Other properties can be also imposed, as those studied in [6, 19, 32, 33, 36].

The concept of overlap as a bivariate aggregation operator was introduced in [4] to measure the degree of overlap of an object in a fuzzy classification system with two classes. This concept has been applied to some interesting situations, in which it is necessary to know the degree of overlap within general classification systems, in particular image segmentation problems as that described in [24] (in which it is necessary to discriminate between object and background) or in the framework of preference relationships [5].

Obviously, there are situations in which we need to measure the degree of overlapping of an object in a fuzzy classification system with more than two classes. Thus, with the aim of so-extending this concept, the concept of an overlap function was generalized into a n-dimensional framework in [18]. Through this generalization, it is possible to analyze most relevant properties and applications. Indeed, in this work we propose an application of n-dimensional overlap and grouping functions to community detection problems into a fuzzy framework.

The definition of an overlap function and some basic results about it were presented in [4, 24]. Particularly, an overlap function is defined as a particular type of bivariate aggregation function characterized by a set of symmetry, natural boundary and monotonicity properties.

Definition 2.1.

$$G_O : \ [0,1]^2 \longrightarrow [0,1]$$

is an overlap function if and only if the following holds:

1. *G_O is symmetric.*
2. *$G_O(x, y) = 0$ if and only if $xy = 0$.*
3. *$G_O(x, y) = 1$ if and only if $x = 1$ and $y = 1$.*
4. *G_O is non-decreasing.*
5. *G_O is continuous.*

Let us observe (as shown in [4, 5, 24]), that overlap functions are closely related with t-norms, but present some differences since the associative property is not required for the former (while it is for the latter). In the following example, we can see an instance of an aggregation function that is an overlap function but not a t-norm if $p > 1$.

Example 2.1. *It is easy to see that the bivariate aggregation function $G_p(x, y) = (\min\{x, y\})^p$ is an overlap function, since the properties (1)-(5) are satisfied. But let us also note that, when $p > 1$, the bivariate function G_p is not associative, and thus it is not a t-norm.*

Let us now recall in the notion of grouping function, also proposed in [4, 24] as a natural complement to overlap functions. Given two degrees of membership $x = \mu_A(c)$ and $y = \mu_B(c)$ of an object c into classes A and B, a grouping function is supposed to yield the degree z up to which the combination (grouping) of the two classes A and B is supported, that is, the degree up to which either A or B (or both) hold.

Definition 2.2. *A grouping function is a function*

$$G_G : [0, 1]^2 \longrightarrow [0, 1]$$

that satisfies the following conditions:

1. *G_G is symmetric.*
2. *$G_G(x, y) = 0$ if and only if $x = y = 0$.*
3. *$G_G(x, y) = 1$ if and only if $x = 1$ and $y = 1$.*
4. *G_G is non-decreasing.*
5. *G_G is continuous.*

2.2. n-dimensional overlap functions

In [18], the previous ideas presented for two sets or classes were extended into a more general case. Sometimes, an object may belong to more than two classes, and thus it may be interesting to measure the degree of overlap of this object with respect to the classification system given by the available classes.

Definition 2.3. *An n-dimensional aggregation function $G_O : [0, 1]^n \longrightarrow [0, 1]$ is an n-dimensional overlap function if and only if:*

1. *G_O is symmetric.*
2. *$G_O(x_1, \ldots, x_n) = 0$ if and only if $\prod_{i=1}^{n} x_i = 0$.*
3. *$G_O(x_1, \ldots, x_n) = 1$ if and only if $x_i = 1$ for all $i \in \{1, \ldots, n\}$.*
4. *G_O is increasing.*
5. *G_O is continuous.*

In a similar way, the grouping concept can be also extended into a more general framework. Given n degrees of membership $x_i = \mu_{C_i}(c)$ for $i = 1, \ldots, n$ of an object c into classes $C_1, \ldots, C_n$, a grouping function is supposed to yield the degree z up to which the combination (grouping) of the n classes $C_1, \ldots, C_n$ is supported.

Definition 2.4. *An n-dimensional function*

$$G_G : [0,1]^n \longrightarrow [0,1]$$

is an n-dimensional grouping function if and only if it satisfies the following conditions:

1. G_G *is symmetric.*
2. $G_G(x) = 0$ *if and only if* $x_i = 0$, *for all* $i = 1, \dots, n$.
3. $G_G(x) = 1$ *if and only if there exist* $i \in \{1, \dots, n\}$ *with* $x_i = 1$.
4. G_G *is non-decreasing.*
5. G_G *is continuous.*

Again, continuous t-conorms (their n-ary forms) and their convex combinations are prototypical examples of n-ary grouping functions.

Example 2.2. *The following aggregation functions are examples of n-dimensional grouping functions:*

- *The maximum powered by p:* $G_G(x_1, \dots, x_n) = \max\limits_{1 \leq i \leq n} \{x_i^p\}$ *with* $p > 0$.
- *The Einstein sum aggregation operator:* $ES(x_1, \dots, x_n) = \dfrac{\sum_{i=1}^n x_i}{1 + \prod_{i=1}^n (x_i)}$

2.3. Community detection problems with overlapping communities

In order to introduce the formal definitions of community detection problems with or without overlapping communities, let us introduce the following notation:

- $V = \{1, 2, \dots, n\}$ will represent the finite set of objects in the network, i.e. the elements to be clustered.
- $E = \big\{ \{i,j\} \mid i, j \in V \big\}$ will be the set of non-ordered pairs of related (neighboring) items of V. In this way, if two elements $i, j \in V$ are related, then there exists an edge $e = \{i, j\} \in E$; otherwise, $\{i, j\} \notin E$. Let m be number of edges ($m = \big| E \big|$).

Hence, we have a graph $G = (V, E)$ that shows the relationships between the items. The graph G can be assumed to be connected; otherwise, its connected components can be analyzed separately.

Classical community detection problems can be viewed as graph partition problems. In this way, the obtained family of clusters can be viewed as the first step towards a posterior segmentation or classification, depending on the final objective of our analysis. Clustering network problems, also addressed as community detection problems in networks, are usually defined as the problem of finding a *good* partition for a given graph $G = (V, E)$.

Definition 2.5. *Given a graph $G = (V, E)$, we will say that the set $\mathcal{C} = \{C_1, \dots, C_r\}$ is a community detection solution (without overlapping communities) if and only if*

- $C_i \cap C_j = \emptyset$ *for all* $i \neq j$ *(non-overlapping communities),*

- $\bigcup_{j=1}^{r} C_j = V$,
- *each subgraph* $(C_i, E_{|C_i})$ *is a connected graph for all* i.

Currently, there exist a huge number of algorithms that address the identification of classical communities (for more details see for example [20], or the exhaustive review of Fortunato [15] or). Clustering networks algorithms detect communities according to topological structures or dynamical behaviors of networks. Depending on the characteristics of the algorithms that cluster the graph into communities or groups, many classifications are possible. For example, Fortunato [15] establishes a division between traditional methods, partitional clustering, spectral clustering, divisive algorithms, modularity-based methods and dynamic algorithms.

As has been pointed out in the introduction, there exist many real situations (see [38]) in which one node should belong to more than one community. Taking into account this, classical community detection problems can be extended to a more general formulation in which what we actually look after is just the covering of the network.

Definition 2.6. *Given a graph* $G = (V, E)$*, we will say that the set* $\mathcal{C} = \{C_1, \ldots, C_r\}$ *is an overlapping community detection solution if and only if*

- $\bigcup_{j=1}^{r} C_j = V$,
- *each subgraph* $(C_i, E_{|C_i})$ *is a connected graph for all* i.

The applications of this problem are diverse. On one hand, as a natural extension of classical community problems, in which more complex and realistic situations are allowed. On the other hand, and related to some classical social network problems, this more complex problem serves the purpose of detecting and identifying key nodes for a given structure (usually referred to as *overlapping nodes*). These nodes play a special role in complex network systems. However, most known classical algorithms, such as divisive algorithm [16] or agglomerative [15], cannot detect them.

Finally, to conclude this section we will also present a formulation of fuzzy community detection problems. Fuzzy community detection problems are usually understood as a natural extension of community detection problems with overlapping communities, in which we search for a *good* fuzzy partition of the set of nodes in the graph. It is clear that once it is allowed that a node can belong to more than one cluster or community, it seems reasonable to extend this idea by considering a fuzzy membership function μ_{C_i} to be associated to each community C_i under consideration. Nevertheless, there are many authors (see for example [23] among others) that define fuzzy community detection problems by imposing the constraint $\sum_{l=1}^{r} \mu_{C_l}(i) = 1$ for any node i in the network, which implies that this fuzzy solution is in fact a Ruspini fuzzy partition [34]. However, it is important to notice that this constraint actually imposes the so-defined fuzzy community detection problems to be closer to the classical community detection problems than to overlapping community detection problems. And even worse, this constraint implies that fuzzy community detection problems are not an extension of overlapping community detection problems. For these reasons, we prefer the following definition of fuzzy community detection problems.

Definition 2.7. *Given a graph* $G = (V, E)$*, we will say that the fuzzy clusters* $\{\widetilde{C_1}, \ldots, \widetilde{C_r}\}$ *over the set* V *with membership functions* $\mu_{C_1}, \ldots \mu_{C_r} : V \longrightarrow [0, 1]$*, are a solution of a fuzzy community detection problem if and only if* $\forall\ i \in V$, $\max\{\mu_{C_l}(i),\ 1 \leq l \leq r\} > 0$.

Consistently with [1, 2], we would like to emphasize that we do not impose that $\sum_{l=1,r} \mu_{C_l}(i) = 1$. Rather, we only need to impose that all nodes in the network belong to at least one class with degree strictly greater than zero. Let us observe that all fuzzy community detection solutions that impose the previous Ruspini-like condition are a particular case of our previous definition, but the opposite is not true. Also, let us observe that such Ruspini-like assumption eliminates all overlapping community detection solutions as possible solutions for a fuzzy community detection problem. Instead, by adopting our definition, the following proposition trivially holds.

Proposition 2.1. *Any classical community detection solution is a particular solution of the community detection problem with overlapping communities. Also, any overlapping community detection solution is particularly a solution of the fuzzy community detection problem.*

3. Modularity measure in fuzzy community detection problems

3.1. State of the art

Once the classical community detection problem, overlapping community detection problem and the fuzzy community detection problem have been formally defined, it is time to measure how good the solutions are. Modularity is one of the most used measures to quantify the quality of a partition in networks when you don't know the a-priori communities. In the case in which the graph is built artificially and/or the real communities are known a priori, different measures to compare two partitions can be found in the literature (see [15]). Nevertheless, in real networks, in which quite often there is not a priori knowledge about the actual communities, modularity is still the best choice to determine if a partition is good or not. For unsupervised clustering algorithms, modularity can also be used to determine the optimal number of communities. Moreover, modularity is often used to compare the performance among several algorithms. This measure was initially defined by Girvan and Newman in [16] for crisp partitions and crisp graphs. In this paper it will be denoted by Q_{GN}. Particularly, given a network $G = (V, E)$ and a partition $\mathcal{C}$ in the conditions of Definitions 2.5 or 2.6, the Girvan-Newman modularity of such a partition is defined as:

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

where m is the number of edges in the graph, k_i is the degree of node i, A_{ij} is the adjacency matrix of the graph and $\delta(c_i \; c_j)$ is equal to 1 if nodes i and j belong to the same cluster and 0 otherwise. The modularity of a partition represents the fraction of edges that fall within the given groups minus the expected such fraction if edges were distributed at random.

Remark 1. Let us observe that this definition allows to measure the performance of a crisp clustering of a graph with overlapping communities (i.e. a node can belong to more than one class).

Although the above modularity measure presents some issues (as for example the resolution limit), currently it is the most used measure to quantify the quality of a solution for community detection problems, with and without overlapping communities, when there is not an a priori known community structure.

Taking into account that there are few methods that produce a fuzzy partition of a network (see for example [38]), few efforts has been dedicated to extend the crisp modularity measure into a more general scenario. Next we give a very short review of the different extensions of the modularity measure in a fuzzy framework (see [15] for more details).

In [40], it is presented one of the first definitions that permit to measure the modularity of a fuzzy partition of a network. In that paper, the authors propose a fuzzy modularity measure based on the α-cuts of such fuzzy partition, in the following sense. Recall that given a network $G = (V, E)$ and a fuzzy partition $\{C_1, \ldots, C_r\}$ of V, the set $V_c = \{i \in V \;/\; \mu_{C_k}(i) \geq \alpha\}$ is a crisp community for all $k \in \{1, \ldots, r\}$ and for any value of $\alpha \in [0, 1]$.

Definition 3.1. *Given a fuzzy partition $\{C_1, \ldots, C_r\}$ of a network $G = (V, E)$ and its corresponding α-cuts V_c for a given $\alpha \in [0, 1]$, Zhang's fuzzy modularity is defined as:*

$$Q_{Zhang}(\alpha) = \sum_{c=1}^{r} [\sum_{i,j \in V_c} \frac{((\mu_{C_c}(i) + \mu_{C_c}(j))/2)\, A_{ij}}{2m} - \left(\sum_{i \in V_c, j \notin V_c} \frac{((\mu_{C_c}(i) + 1 - \mu_{C_c}(j))/2)\, A_{ij}}{2m} \right)^2].$$

The previous definition presents some problems since the modularity of a fuzzy partition depends on the value of α, that is for each $\alpha \in [0, 1]$ a different value of the modularity measure is obtained. In [27], an alternative definition of fuzzy modularity was presented by Liu, which not depends on a parameter α.

Definition 3.2. *Given a fuzzy partition $\{C_1, \ldots, C_r\}$ of a network $G = (V, E)$, let now the crisp communities V_c be defined as $V_c = \{i \in V \;/\; \mu_{C_c}(i) = max\{\mu_{C_k}(i) \; 1 \leq k \leq r\}$. Then, Liu's fuzzy modularity is defined as:*

$$Q_{Liu} = \sum_{c=1}^{r} \left[\sum_{i,j \in V_c} \frac{((\mu_{C_c}(i) + \mu_{C_c}(j))/2)\, A_{ij}}{2m} - \left(\sum_{i \in V_c, j \notin V_c} \frac{((\mu_{C_c}(i) + 1 - \mu_{C_c}(j))/2)\, A_{ij}}{2m} \right)^2 \right]$$

Obviously, the main difference between Q_{Liu} and Q_{Zhang} lies on the definition of the crisp communities V_c. Let us observe that if the fuzzy partition $\{C_1, \ldots, C_r\}$ is a Ruspini partition (i.e. for all $i \in V$, $\sum_{c=1,\ldots,r} \mu_{C_c}(i) = 1$), then it is possible to find values of α for which both measures coincides.

The third well-known generalization of the crisp modularity measure in a fuzzy framework is given in [29]. In that paper, the Kronecker delta $\delta(c_i \; c_j)$ that appears in the classical formula given by expression (1), is replaced by s_{ij}, where s_{ij} denotes the sum of the products of the membership degrees of nodes i and j in the communities to which they both belong. Formally, the fuzzy modularity measure introduced in [29] by Nepusz et al. can be expressed as follows:

$$Q_{NE} = \frac{1}{2m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i \; k_j}{2m} \right] s_{ij}, \tag{2}$$

where $s_{ij} = \sum_{c=1,\ldots,r} \mu_{C_c}(i)\mu_{C_c}(j)$. In the previous formula it is imposed that the fuzzy partition $\{C_1, \ldots, C_r\}$ of the graph has to be a Ruspini partition, in the sense that $\sum_{c=1,\ldots,r} \mu_{C_c}(i) = 1$ for any $i \in V$. By imposing the Ruspini condition it is guaranteed that all the s_{ij} belong to the unit interval $[0, 1]$.

In the following example we present the result of calculating these three fuzzy measures modularity from a fuzzy clustering of a well-known network.

Example 3.1. *In [40], it is introduced a network in which there exist clearly overlapping nodes (see Figure 1). This situation presents in a natural way 3 overlapping communities with some nodes as* 5, 9, 2 *or* 13 *(especially* 5 *and* 9*) that may belong to more than one community.*

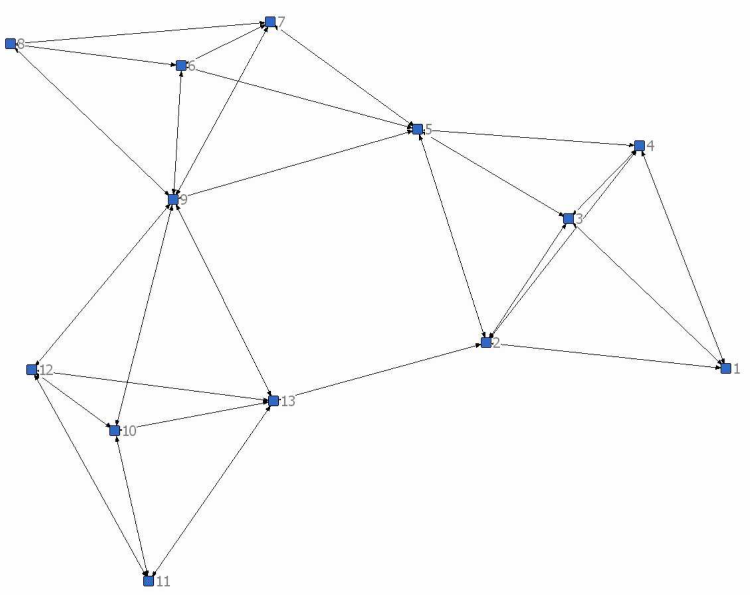

Figure 1: Zhang's network with overlapping communities.

In Table 1, we show a fuzzy clustering of the 13 *nodes in three communities.*

Table 1: A fuzzy clustering of *the Zhang network.*

Nodes	$\mu_{C_1}(i)$	$\mu_{C_2}(i)$	$\mu_{C_3}(i)$
1	0	0	0.9898
2	0.0152	0.0189	0.9898
3	0.0017	0	0.9901
4	0.0017	0	0.9901
5	0.9774	0	0.9728
6	0.9900	0	0
7	0.9900	0	0
8	0.9898	0	0
9	0.9898	0.3463	0
10	0.0017	0.9901	0
11	0	0.9898	0
12	0.0017	0.9901	0
13	0.0173	0.9897	0.0210

Given the fuzzy clustering $\mu = (\mu_{C_1}, \mu_{C_2}, \mu_{C_3})$ shown in Table 1, in Table 2 we show the values attained by the modularity measures $Q_{Zhang}(\alpha)$, Q_{Liu} and Q_{NE} for different thresholds α . As it can be observed, both Q_{Liu} and Q_{NE} modularity measures do not depend on α. Also let us note that the modularity Q_{NE} assumes a probabilistic scenario and requires the fuzzy clustering to be a Ruspini partition. Taking into account that μ is not a fuzzy partition in the sense of Ruspini, Q_{NE} does not perform well. Classical Girvan-Newman modularity cannot be computed since the detected communities are not crisply defined.

Table 2: Modularity measures in terms of α .

α	$Q_{Zhang}(\alpha)$	Q_{Liu}	Q_{NE}
$(0, 0.0017]$	-0.527	0.4215	1.21
$(0.0017, 0.0173]$	-0.0714	0.4215	1.21
$(0.0173, 0.021]$	0.2470	0.4215	1.21
$(0.021, 0.3463]$	0.3307	0.4215	1.21
$(0.3463, 0.9728]$	0.3968	0.4215	1.21
$(0.9728, 0.9774]$	0.4215	0.4215	1.21
$(0.9774, 1]$	Not defined	0.4215	1.21

4. A new modularity measure for fuzzy community detection outputs

Although the fuzzy modularity presented in [29] is close to the correct generalization of the classical modularity measure Q_{GN} into a fuzzy scenario, it presents some deficiencies. The most important one is that it is necessary to impose that the fuzzy partition of the set of nodes has to be a Ruspini partition, and thus is not a generalization of the classical modularity measure when there exist overlapping communities. Therefore, if we have a crisp or fuzzy partition in which for one node i it is $\sum_{c=1}^{r} \mu_{C_c}(i) > 1$, then Q_{NE} does not perform well.

In the original definition of Girvan-Newman modularity, $\delta(c_i\ c_j)$ represents the truth-value associated with the assertion *node i and node j belong to the same community.* In Q_{NE} this degree of truth of *node i and node j belong to the same community* is replaced by $s_{ij} = \sum_{c=1}^{r} \mu_{C_c}(i)\mu_{C_c}(j)$, which in fact exhibits a different meaning. As it is pointed in [23], the modularity measure Q_{NE} does not permit overlapping in the sense that $\sum_{c=1}^{r} \mu_{C_c}(i) > 1$ (either in the crisp or fuzzy case), and thus it is not a generalization of the crisp GN modularity measure with overlapping nodes.

Now let us try to quantify the assertion *node i and node j belong to the same community* by means of overlap and grouping functions. Let us observe that the $\delta(c_i\ c_j)$ in the classical, crisp modularity measure Q_{GN} takes value 1 if and only if both i and j belong to the same community, for any of the communities in the network partition $\{C_1, \ldots, C_r\}$. If we denote by $C^i = \cup_{l\ /\ i \in C_l} C_l$ and $C^j = \cup_{l\ /\ j \in C_l} C_l$ the sets of communities to which nodes i and j respectively belong, then $C^i \cap C^j = \cup_{l\ /1, j \in C_l} C_l$ represents the set of communities in which i and j belong simultaneously. In a crisp scenario, $\delta(c_i\ c_j) = 1$ if and only if $\cup_{l\ /\ i,j \in C_l} C_l \neq \emptyset$. In a fuzzy framework, this union can be represented by means of a grouping function G_G and the intersection or the condition of both $i, j \in C_l$ through an overlap function.

To illustrate this idea, let i and j be two nodes such that their membership functions to the three communities C_1, C_2 and C_3 are respectively given by $\mu(i) = (\mu_{C_1}(i), \mu_{C_2}(i), \mu_{C_3}(i)) = (0.9, 1, 0)$ and $\mu(j) = (\mu_{C_1}(j), \mu_{C_2}(j), \mu_{C_3}(j)) = (0.4, 0.5, 1)$. The truth-degree of the assertion *nodes i and j belong simultaneously to the community* C_1 could be measured as the degree of overlap that this community has over the nodes i and j, i.e. $G_O(\mu_{C_1}(i), \mu_{C_1}(j)) = G_O(0.9, 0.4)$. And then, after obtaining the degrees up to which i and j belong to communities C_1, C_2 and C_3, it is possible to aggregate these three values into a single one by using a grouping function. Thus, we propose to redefine the s_{ij} as

$$s_{ij} = G_G\left(G_O(\mu_{C_1}(i), \mu_{C_1}(j)), \ldots, G_O(\mu_{C_3}(i), \mu_{C_3}(j))\right) \tag{3}$$

For example, by taking the overlap function $G_O(x, y) = \min\{x, y\}^{1/2}$ and the grouping function $G_G(x_1, \ldots, x_n) = \max\{x_i, \}$, we obtain that $s_{ij} = \max\{0.4^{1/2}, 0.5^{1/2}, 0\} = 0.5^{1/2} = 0.707$.

Taking into account the previous considerations, in the next definition we present an extension of the classical modularity measure Q_{GN} enabling to evaluate the performance of a fuzzy classification (not necessarily a Ruspini partition) of the set of nodes of a graph based on grouping and overlap functions.

Definition 4.1. *Given a fuzzy clustering or partition $\mathcal{C}$ of a graph (V, E) with membership functions $\mu_{C_c} : V \longrightarrow [0, 1]$, for all $c \in \mathcal{C}$, its modularity measure is defined as:*

$$\widetilde{Q}(\mathcal{C}) = \frac{1}{2m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i\ k_j}{2m}\right] G_G\left\{G_O\left(\mu_{C_c}(i), \mu_{C_c}(j)\right)\ c \in \mathcal{C}\right\}, \tag{4}$$

where A_{ij} is the adjacency matrix of the crisp graph, m is the number of links of this graph, k_i is the degree of node i in the graph, $G_G : [0, 1]^{|\mathcal{C}|} \longrightarrow [0, 1]$ is an n-dimensional grouping function and $G_O : [0, 1]^2 \longrightarrow [0, 1]$is a bivariate overlap function.

Proposition 4.1. *Given a fuzzy clustering $\mathcal{C}$ of a graph (V, E) with membership functions μ_{C_c} : $V \longrightarrow [0,1]$, for all $C \in \mathcal{C}$, if the fuzzy clustering is a solution of the classical community detection problem (i.e. for all $i \in V, \mu_C \in \{0,1\}$ and $\sum_{C\in\mathcal{C}} \mu_C(i) = 1 \ \ \forall\ i \in V$), then the following holds:*

$$Q_{GN} = \widetilde{Q}(\mathcal{C}).$$

Proof. From the definition of grouping and overlap functions it is clear that given any $i, j \in V$, the expression $G_G \{G_O (\mu_C(i), \mu_C(j)) \ \ C \in \mathcal{C}\}$ coincides with δ_{ij} if the fuzzy clustering is a classical partition.

Proposition 4.2. *Given a fuzzy clustering $\mathcal{C}$ of a graph (V, E) with membership functions μ_C : $V \longrightarrow [0,1]$, for all $C \in \mathcal{C}$, if the fuzzy clustering is a solution of the overlapping community detection problem (i.e. for all $i \in V, \mu_C \in \{0,1\}$) then the following holds:*

$$Q_{GN} = \widetilde{Q}(\mathcal{C}).$$

Proof. From the definition of grouping and overlap functions it is again clear that given any $i, j \in V$, the expression $G_G \{G_O (\mu_C(i), \mu_C(j)) \ \ C \in \mathcal{C}\}$ coincides with δ_{ij} if the fuzzy clustering is a classical covering.

Let us note that Proposition 4.1 is also satisfied by measures Q_{NE}, Q_{Liu} and Q_{Zhang}. That is, these three measures coincide with the Girvan-Newman modularity when facing crisp, non-overlapping communities. However, Proposition 4.2 does not hold for any of these three measures, that is, they do not coincide with Girvan-Newman modularity in case the crisp communities present overlapping. This is an important feature of the modularity measure $\widetilde{Q}(\mathcal{C})$ here proposed: it coincides with Girvan-Newman modularity both in the overlapping and non-overlapping scenarios.

In Table 3, we summarize these results. In the first column, it is said 'yes' if the measure is able to deal with classical crisp (non-overlapping) partitions; in the second column, 'yes' means that the measure is able to deal with overlapping community detection solutions (or crisp coverings); the third column evaluates if a measure is able to deal with fuzzy community detection solutions assumed to be Ruspini partitions (i.e. fuzzy partitions), while this assumption is removed for the fourth column (that instead refers to fuzzy clusters rather than partitions); and finally, the last column assesses whether the measure coincides with the classical Girvan-Newman modularity Q_{GN} for crisp (non-overlapping) partitions.

Table 3: Different modularity measures.

Measures	Crisp Partition	Crisp Covering	Fuzzy Partition	Fuzzy Clustering	Extension
GN Modularity	YES	YES	NO	NO	-
Zhang Modularity	YES	NO	YES	NO	YES
Liu Modularity	YES	NO	YES	NO	YES
Nepusz Modularity	YES	NO	YES	NO	YES
New Modularity	YES	YES	YES	YES	YES

Thus, to sum up, let us remark that if the partition $\mathcal{C}$ is crisp, then $\widetilde{Q}$ coincides with the classical modularity measure Q_{GN} (despite overlapping communities being allowed or not, as explained above). Let us also observe that our definition allows measuring the performance of a crisp overlapping classification as well as of a fuzzy overlapping classification in the sense that $\sum_{C\in\mathcal{C}} \mu_C(i) > 1$. It can be easily proved that our fuzzy modularity measure $\widetilde{Q}$ is a generalization of the classical Q_{GN} when there exist overlapping communities. However, it is important to remark that neither Q_{Zhang}, Q_{Liu} nor Q_{NE} can provide a suitable measure in the case of a crisp partition with overlapping communities. Therefore, let us stress the relevance of the proposed fuzzy modularity measure $\widetilde{Q}$, since it allows to measure the performance of a fuzzy network clustering, naturally extending the classical modularity measure for crisp overlapping communities to the fuzzy case.

5. Identifying fuzzy communities, overlapping nodes and ranking them

In the previous sections we have been addressing three different problems or detection tasks: the classical community detection (CCD for short) problem, the overlapping community detection (OCD) problem and the fuzzy community detection (FCD) problem. Together with these three tasks or problems, in this section we now also focus on two other problems: the overlapping node detection (OND) problem, and the problem of ranking the overlapping nodes (RON).

In the first, OND problem, the main aim is to classify the nodes into two groups: overlapping nodes (that are those nodes that acts as intermediaries between two or more communities) and those others for which this role is not relevant. Several algorithms have been proposed in the literature (see [22] for example) to deal with the OND problem. The second problem, RON, is related with the study of measures that quantify the intermediation power of nodes in a community detection structure. Although this problem could seem similar to the definition of a betweeness centrality measure, the difference between them is that the inherent community structure is not taken into account in a betweeness centrality measure. Betweeness centrality measures (see e.g. [17] for more details) try to capture the intermediation power of one node in the communications between the rest of the nodes in the network, but not between communities. To the extent of our knowledge, the RON problem has not been formally studied in the related literature.

Let us note that the five problems or tasks above referred are strongly related, since the following six implications between them hold (we use the inclusion symbol $\subset$ to denote that some problems are particular instances of other problems, and the implication symbol $\longrightarrow$ to denote that solutions to some problems can be obtained from those of other problems) :

- $CCD \subset OCD$. Any solution to a classical community detection problem is also a solution to the overlapping community detection problem (first part of Proposition 2.1).

- $OCD \subset FCD$. Any classical overlapping community detection solution is a fuzzy community detection solution (second part of Proposition 2.1).

- $OCD \longrightarrow OND$. From any solution to the (crisp) overlapping community detection problem, an overlapping node detection solution can be derived in a trivial way: if a node belongs to more than one community then this node is an overlapping node. Otherwise, the node is not an overlapping node.

- $FCD \longrightarrow OCD \longrightarrow OND$. From any fuzzy community detection solution, we may obtain (for example by fixing an alpha-cut) an overlapping community detection solution. From this last, attending to the previous item, an overlapping node detection solution can be easily obtained. Nevertheless, the implication FCD to OCD can be addressed in different ways and is not necessarily a trivial task.

- $RON \longrightarrow OND$. From any ranking of the overlapping nodes (i.e. from any solution to RON), it is possible to obtain a solution to the overlapping node detection problem.

- $FCD \longrightarrow RON$. Given a solution to the fuzzy community detection problem, we could derive a degree of overlap for each node, which can be used to construct a ranking of the overlapping nodes, i.e. a solution to RON. Nevertheless, this implication is not necessarily a trivial task.

Many algorithms have been proposed for the classical community detection (see [15, 20] for more details) and the overlapping community detection problems (see [22, 40]). Nevertheless, very few methods have been proposed that deal with the fuzzy community detection problem. Moreover, up to our knowledge, no method have been proposed that deals with the FCD, OND and RON problems simultaneously. Based on this observation, and taking into account that the modularity measure proposed in this paper is able to deal with any of the CCD, OCD and FCD problems, we present a general method that finds a consistent solution to the FCD, OCD, OND and RON problems. In the following subsections we explain its steps in detail.

5.1. Identifying Fuzzy Communities

As previously pointed out, very few algorithms have been proposed that produce a fuzzy clustering of a network. In [29], a fuzzy clustering is obtained through a non-linear constrained optimization problem solved as a quadratic-complexity algorithm. In [40], the network is transformed into a $k-1$-dimensional Euclidean space and then the fuzzy c-means (FCM) algorithm is used to detect up to k communities. Other algorithms as the FOG algorithm [14] or the NMF algorithm [31] present a probabilistic solution (i.e. a probabilistic classification of the nodes) that cannot be considered as a fuzzy algorithm in the usual sense. Later, in [41] a similar method to [40] is provided, in which the network structure is mapped into a low-dimensional space through a multidimensional scaling (MDS) approach. After that, FCM is employed to find fuzzy communities in the network. The number of communities is determined by means of the fuzzy modularity measure Q_{NE} proposed in [29]. In this paper we have been applied the FCM algorithm without normalization.

Although the main aim of this paper is to present a new modularity measure that fixes the main deficiencies of the preexisting ones, it is possible to rebuild some of the few algorithms that actually produce a fuzzy clustering of a network by using this new modularity measure. With this aim, in this section we provide a new fuzzy community detection algorithm based on that proposed in [40], but taking into account the new modularity measure here defined to determine the optimal number of classes.

Fuzzy Community Detection Method: ($FCD_{z-grouping}$)

- For each possible number of communities $c \in \{2, \dots, n\}$:

  (1) Obtain a fuzzy clustering of the network $\mu_1, \dots, \mu_c$ with c classes (e.g. by means of the Zhang or the Wang algorithms).

  (2) Compute the fuzzy modularity function of the previous fuzzy partition $\widetilde{Q}(\mu)$.

- Pick the number of classes c and the corresponding fuzzy partition μ that maximize the modularity function $\widetilde{Q}(\mu)$.

Example 5.1. The Zhang's example. *Now we come back to the previously presented example given in [40]. As already discussed, this network (shown in Figure 1) clearly presents three overlapping communities, with some nodes as* 5*,* 9*,* 2 *or* 13 *(especially* 5 *and* 9*) that could belong to more than one community.*

Table 4 shows the obtained modularity measure for each number of communities, when using the composition of overlap and grouping functions $max\{i \neq jMin^{1/2}\{x_i, x_j\}\}$ to compute the fuzzy deltas s_{ij} . Let us observe that the optimal number of communities is 3. The obtained fuzzy community detection solution is the same that was presented in Table 1 of Example 3.1. As can be observed the optimal number of communities is 3.

Table 4: Modularity versus number of communities in the $FCD_{z-grouping}$ method.

Number of Communities	Modularity
2	0.2927
3	**0.3699**
4	0.2288
5	0.1250
6	0.0610
7	0.0215
8	0.0005
9	0
10	0
11	0
12	0
13	0

5.2. Overlapping nodes detection based on fuzzy clustering

As previously mentioned, one of the most important problems in community detection problems is the identification of the overlapping nodes of a network (OND problem). In this section we will describe how to deal with the OND problem based on the $FCD_{z-grouping}$ method previously defined. Let us note that the fuzzy community detection algorithm used in this work is the Zhang's method described in [40], but any other algorithm able to cluster a network may be used in a similar way. A more-in-deep analysis of the interactions between different FCD algorithms with the modularity measure here proposed could be an interesting issue for future work, as it lies outside the scope of this work.

Now we present a simple method to produce a solution of the overlapping community detection (OCD) problem, and thus also of the overlapping node detection (OND) problem.

$OCD_{z-grouping}$ **method:**

- For each possible number of communities $c \in \{2, \dots, n\}$:

  (1) Obtain a fuzzy clustering of the network $\mu_1, \dots, \mu_c$ with c classes (e.g. by means of the Zhang or the Wang algorithms).

  (2) Determine the value of α for which the crisp covering C obtained from the previous fuzzy clusters through the corresponding α-cut maximizes $\widetilde{Q}(C)$. Let us denote this optimal modularity by $\widetilde{Q}_c$.

- Pick the number of classes c and the corresponding crisp covering C that provides the maximum modularity $\widetilde{Q}_c$.

- From the (crisp) overlapping community detection solution determined in the previous step, obtain the set of overlapping nodes.

Let us observe that the previous method produces as an output a fuzzy clustering of the network and also a crisp clustering with overlapping communities in which the overlapping nodes are identified in a crisp way. Also let us note that this method depends on the algorithm by which the fuzzy clustering of the network is obtained for each fixed number of classes c (step 1).

Remark 2. Let us stress that two parameters has to be determined in the previous method: the number of communities c and the value of α that provides the best crisp partition. Thus, the $OCD - z$ method has to solve a bi-level optimization problem in order to attain the overlapping communities with optimal modularity. This procedure may be computationally expensive in dense networks with a big number of nodes. A less computationally expensive alternative is to adopt a lexicographic approach to approximate the solution of the previous bi-level optimization problem, first obtaining the optimal number of fuzzy communities c, and then the optimal α-cut for such c. In this way, the bi-level problem is transformed in a sequence of two single-level problems. Obviously, better results are usually obtained by simultaneously optimizing both parameters of the bi-level problem, although the significance of such improvement may not always compensate for the higher computational costs derived of the bi-level optimization. Particularly, we have observed that in practice both methods usually provide solutions with a similar number of clusters. To illustrate both approaches (bi-level and lexicographic), in the following example we apply this lexicographic alternative, while the bi-level approach will be used in the computational experiments described in the next section.

Example 5.2. Identifying overlapping nodes and communities in the Zhang's example. *Following with the example of the Zhang's network, in Example 5.1 we reached to the conclusion that* $c = 3$ *is the optimal number of communities* ($\widetilde{Q}(C) = 0.37$*, as shown in Table 4), leading to the corresponding fuzzy classification of the* 13 *nodes given in Table 1. From this last table, it is possible to detect that nodes* 5*,* 9*,* 2 *or* 13 *are the most overlapping ones and may belong to more than one community. However, together with this last classification, Table 5 also shows the solution to the overlapping community detection problem obtained for* $\alpha = 0.3$*, which for* $c = 3$ *leads to the maximum modularity* $Q_{GN}(C) = \widetilde{Q}(C) = 0.4847$*. This modularity is associated to the (crisp) partition* $C = \{C_1, C_2, C_3\}$*, where* $C_1 = \{5, 6, 7, 8, 9\}$*,* $C_2 = \{9, 10, 11, 12, 13\}$ *and* $C3 = \{1, 2, 3, 4, 5\}$*. As a consequence, only nodes* 5 *and* 9 *are in this way identified as the overlapping nodes of this network, since no other nodes are being assigned to more than one community. Let us observe that most of the algorithms for (crisp) overlapping node detection as CONGA [22], NMF [31] or CFINDER [30] with* $k = 3, 4$ *coincide with the previous results of our method, both in the crisp clustering* C *as well as in the detected overlapping nodes.*

Table 5: Fuzzy classification using our algorithm in the Zhang's network. In black color the overlapping nodes.

Nodes	$\mu_{C_1}(i)$	$\mu_{C_2}(i)$	$\mu_{C_3}(i)$	Crisp C
1	0	0	0.9898	(0,0,1)
2	0.0152	0.0189	0.9898	(0,0,1)
3	0.0017	0	0.9901	(0,0,1)
4	0.0017	0	0.9901	(0,0,1)
5	0.9774	0	0.9728	**(1,0,1)**
6	0.99	0	0	(1,0,0)
7	0.99	0	0	(1,0,0)
8	0.9898	0	0	(1,0,0)
9	0.9898	0.3463	0	**(1,1,0)**
10	0.0017	0.9901	0	(0,1,0)
11	0	0.9898	0	(0,1,0)
12	0.0017	0.9901	0	(0,1,0)
13	0.0173	0.9897	0.021	(0,1,0)

5.3. Ranking based on overlapping degree

In network analysis, the identification of intermediate nodes (overlapping nodes or bridge nodes) is an important topic, since they have the power of intermediating between the groups to which they belong. The identification of overlapping nodes has been useful in practical applications in different areas as biology (overlapping nodes are the key nodes in proteins interaction networks), communication networks (it has been proved that overlapping nodes spread information in a faster way that other non-overlapping nodes, see for example viral marking problems or the general information diffusion topic), disease spreading, transport problems or security problems among many other disciplines [13, 26]. The practical relevance of overlapping nodes has led to the introduction of the OND problem addressed above. Nevertheless, the OND problem classifies the set of nodes just as overlapping or non-overlapping nodes, but there exist many situations in which it could be useful to have a ranking of (all) the nodes based on its intermediation power. Betweeness centrality (or in general centrality measures) is commonly used for this kind of practical application as it allows ranking the nodes from the most influential ones to the least influential ones. But in a network in which a community structure has been identified such an intermediation power may be better captured in terms of an $overlapping - ness$ measure (since information spread with a higher probability among members of the same community), quantifying the degree up to which a node is an overlapping one. This motivates the problem of constructing a ranking of the nodes based on such an $overlapping - ness$ measure, i.e. the RON problem.

In this section, we will discuss how to obtain such an overlapping degree in order to rank the nodes, from the most overlapping to the least. Given a fuzzy classification of the nodes, that is, a solution of FCD, the simplest situation is that where we want to quantify the degree up to which a node belongs to just two communities. Given two communities C_r and C_s, in order to obtain the degree up to which a node i belong to both communities we should aggregate the values $\mu_{C_r}(i)$ and $\mu_{C_s}(i)$. It seems logical to aggregate them by means of an overlap function, representing the extent up to which node i simultaneously belongs to both communities, i.e. $G_O\left(\mu_{C_r}(i), \mu_{C_s}(i)\right)$. However, a node can be intermediary of any pair of communities C_r and C_s, so we again should aggregate the value $G_O\left(\mu_{C_r}(i), \mu_{C_s}(i)\right)$ for all pairs $r \neq s$. As it is logical to expect that a node has a high degree of bivariate overlapping if he belongs to at least two communities with a high degree, then we can use a grouping function G_G to aggregate these $s = \begin{pmatrix} c \\ 2 \end{pmatrix}$ values. Formally, we present the following definition.

Definition 5.1. *Given a network $G = (V, E)$ and a fuzzy clustering of the network $\mu_{C_1}, \ldots, \mu_{C_c} : V \longrightarrow [0, 1]$, we define the bivariate overlapping degree of node i as:*

$$2 - overlap_\mu(i) = G_G\left(\ G_O\left(\mu_{C_r}(i), \mu_{C_s}(i)\right); 1 \leq r < s \leq c\right),$$

where $G_G : [0, 1]^s \longrightarrow [0, 1]$ is a grouping function, and $G_O : [0, 1]^2 \longrightarrow [0, 1]$ is a bivariate overlap function.

Let us note that the previous definition only takes into account the bivariate intersection between communities. In this sense, a node that belongs with degree 0.3 to three communities and 0 for the remainder ones should have a low $2 - overlap_\mu(i)$. Nevertheless, it could be the most overlapping node that simultaneously belongs to three (or more) communities. In other words, this node could have an important role from a betweeness point of view, but this fact is not detected by the $2 - Overlaping$ degree measure just defined. Taking into account this, the following definition is proposed:

Definition 5.2. *Given a network $G = (V, E)$ and a fuzzy clustering of the network $\mu_{C_1}, \ldots, \mu_{C_c} : V \longrightarrow [0, 1]$, we define the k-variate overlapping degree of node i as:*

$$k - overlap_\mu(i) = G_G\left(G_O\left(\mu_{C_{r_1}}(i), \ldots, \mu_{C_{r_k}}(i)\right); 1 \leq r_1 < r_2 < \ldots < r_k \leq c\right),$$

where $G_G : [0, 1]^s \longrightarrow [0, 1]$ is a grouping function, with $s = \begin{pmatrix} c \\ k \end{pmatrix}$, and $G_O : [0, 1]^k \longrightarrow [0, 1]$ is a k-dimensional overlap function.

Let us note that for a given number c of classes, the k-overlapping degree is obtained after the evaluation (by means of a grouping function) of all possible groups of k communities ($k < c$). This is the reason why the grouping function ranges from $[0,1]^{\begin{pmatrix} c \\ k \end{pmatrix}}$ to [0,1]. Given k communities $C_1, \ldots, C_k$, the degree up to which a node simultaneously belongs to these k communities is obtained by means of an overlap function. For example, if the number of communities c is 5, and we want to obtain the 3-overlapping degree ($k = 3$) of a node i, we have to evaluate the membership of node i to the $10 = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$ different possible groups of 3 communities of the network. For each of these 10 possible groups, the degree up to which node i simultaneously belongs to the 3 communities in the group is obtained through a 3-dimensional overlap function (ranging from $[0, 1]^3$ to $[0, 1]$). After that, these 10 degrees of membership are aggregated by means of a grouping function.

Example 5.3. Ranking overlapping nodes in the Zhang's example. *Let us once more illustrate the previous notions through the Zhang's example. Recall that in Examples 5.1 and 5.2 we concluded that $c = 3$ is the optimal number of communities, and that for this number of communities,* *the crisp partition producing the maximum modularity has two (bivariate) overlapping nodes,* 5 *and* 9*. In this situation, now we could be interested in assessing whether 5 or 9 is the most overlapping node. Moreover, we may be also interested in detecting those nodes acting as intermediaries between more than two communities (i.e. between the* $c = 3$ *communities), and rank them similarly to the bivariate case.*

Table 6 extends the results of the previous examples by also showing the degrees of 2-overlap and 3-overlap of each node. These last have been obtained by using the squared-root minimum as the overlap function G_O*, and the maximum as the grouping function* G_G*. Notice that node 5 leads node 9 as the most bivariate overlapping node. Many other nodes, as nodes 2 and 13, also present a positive 2-overlap, although they are scarcely significant in this role compared to nodes 5 and 9, the only two that were identified as (crisp) overlapping nodes in Example 5.2. However, it is interesting to note that on the other hand nodes 5 and 9 have degree 0 of 3-overlap (since they do not belong at all to communities 2 and 3, respectively), whereas nodes 2 and 13 present a positive (though small) 3-overlap, as a consequence of belonging to all the three communities with a degree greater than 0. Moreover, it is also possible to suggest that node 13 is slightly more important as a 3-overlapping node than node 2.*

Table 6: Ranking the nodes from our fuzzy classification in the Zhang's network.

Nodes	$\mu_{C_1}(i)$	$\mu_{C_2}(i)$	$\mu_{C_3}(i)$	Crisp C	2-overlap	3-overlap
1	0	0	0.9898	(0,0,1)	0	0
2	0.0152	0.0189	0.9898	(0,0,1)	0.1341	0.1233
3	0.0017	0	0.9901	(0,0,1)	0.0412	0
4	0.0017	0	0.9901	(0,0,1)	0.0412	0
5	0.9774	0	0.9728	**(1,0,1)**	0.9863	0
6	0.99	0	0	(1,0,0)	0	0
7	0.99	0	0	(1,0,0)	0	0
8	0.9898	0	0	(1,0,0)	0	0
9	0.9898	0.3463	0	**(1,1,0)**	0.5885	0
10	0.0017	0.9901	0	(0,1,0)	0.0412	0
11	0	0.9898	0	(0,1,0)	0	0
12	0.0017	0.9901	0	(0,1,0)	0.0412	0
13	0.0173	0.9897	0.021	(0,1,0)	0.1449	0.1315

6. Computational results and final comments

Now, in order to test the effectiveness of the method proposed in the previous section for those three problems, we will compare its performance with that of some widely used algorithms in two well-known networks:

(i) The *Karate Club* network,

(ii) the *Les Miserables* network.

Particularly, for this comparison we have chosen some classical, crisp (non-overlapping) community detection algorithms as *GN*, *CNM* and *D&L* (see [20]), some other algorithms that allow overlapping communities as *NFM* and *CFINDER*, and of course the proposed method in its fuzzy (*FCD-z*) and overlapping (*OCD-z*) versions. Let us recall that all these methods (except $CFINDER$) produce a dendogram on the set of nodes, in such a way that each level of the dendogram identifies a partition of the network with a different number of groups or communities. For each method and each number of groups, the obtained partitions μ will be compared in terms of the proposed modularity measure $\widetilde{Q}(\mu)$. The squared root minimum and the maximum are respectively the overlap and grouping functions used to compute $\widetilde{Q}$.

Let us remark again that $\widetilde{Q}$ coincides with the classical Girven-Newman modularity Q_{GN} both for classical community detection and overlapping community detection problems, as stated by Propositions 4.1 and 4.2. Therefore, as it is devised to deal with different scenarios (classical communities, overlapping communities, fuzzy communities), the proposed modularity measure $\widetilde{Q}$ enables to compare all the chosen methods through the same modularity measure, that behaves as Q_{GN} outside the fuzzy framework of the FCD-z method.

6.1. The Karate Club network

One of the most well-known examples in the literature on social networks or community detection problems is the Karate Club network defined by [39].

In this network, the nodes represent the members of a karate club and the edges represent the friendship relationships between them. The network consists of the 34 members of a karate club as nodes and 78 edges representing their friendship relations, as observed over a period of two years. Due to a disagreement between the administrator and the instructor (nodes 1 and 34), the club splits into two smaller clubs. The question we are concerned with is whether we can uncover the potential behavior of the network, detecting the different communities or groups in which it would split, and particularly identifying to which community each node will belongs to. From an overlapping point of view, an interesting question is to discover the nodes that will belong to more than one community in case the friendship network breaks.

Table 7 shows the performance of the chosen algorithms in terms of $\widetilde{Q}(\mu)$, for the partitions μ obtained by each method for different numbers of communities. The best result achieved by each method is shown in bold. Notice that the best results are obtained with the $OCD - z$ algorithm proposed in this paper, for which it is obtained a covering of the set of nodes with 4 communities achieving a modularity $\widetilde{Q} = Q_{GN} = 0.4362$. Let us note that this modularity is clearly better than that of any solution found by the classical community detection methods, in which overlapping is not allowed.

As it can be observed, the CFINDER algorithm only gives information for 15 groups. This happens because CFINDER is actually an overlapping node detection (OND) algorithm, the output of which only considers the communities with overlapping nodes. Therefore, when the output of this algorithm is adapted to an OCD solution (in which all nodes must belong to some community), the nodes that does not belong to an overlapping community are assumed to be isolated (which is an unrealistic hypothesis). As a consequence of this assumption, this algorithm for overlapping community detection problems usually reaches the optimum at a quite big number of communities, with relatively bad values of modularity. Nevertheless, it is an interesting algorithm to detect overlapping nodes.

Let us also observe that the optimal number of communities differs between the $FCD - z$ and the $OCD - z$ methods, being attained at 3 and 4 communities respectively. This shows that, as discussed in Section 5, the lexicographic version of $OCD - z$ may produce worse results than the bi-level version. In order to present the solution with a greater modularity (i.e. that of the $OCD-z$ method), we have chosen to describe the results of both methods ($FCD - z$ and $OCD - z$) only for the case of 4 communities. Indeed, Table 8 shows the fuzzy clustering solution given by the $FCD-z$ method when we choose four communities to break the network, together with the optimal overlapping communities solution obtained through $OCD - z$ for $\alpha = 0.3$.

Finally, Table 9 shows the obtained 2-overlap and 3-overlap degrees based on the 4-communities fuzzy clustering of the network. Again, the squared root minimum and the maximum are respectively the overlapping and the grouping functions used to compute the k-overlap degrees. As it can be observed, it is possible to rank the nodes based on their overlapping role. Let us note that the top so-ranked nodes in Table 9 coincide with the overlapping nodes detected by standard OND algorithms. In terms of the 2-overlapping degree, the top 5-nodes (from highest to lowest) given by our RON procedure are nodes 29, 24, 10, 9 and 31. The overlapping nodes detected by our OND algorithm are the first 4 nodes of these 5. The NFM or CFINDER algorithms identify as overlapping the nodes 29, 24, 9 and 29, 24, 10, 31 respectively, but with lower modularity values. In general, the results given by our ranking procedure are consistent with the solutions given by classical algorithms as $CFINDER$, $CONGA$ or NFM, but also permit to make a deeper analysis (as our RON solution allows ranking the detected overlapping nodes in terms of their $overlapping - ness$, and to refer this notion to different numbers of overlapping communities).

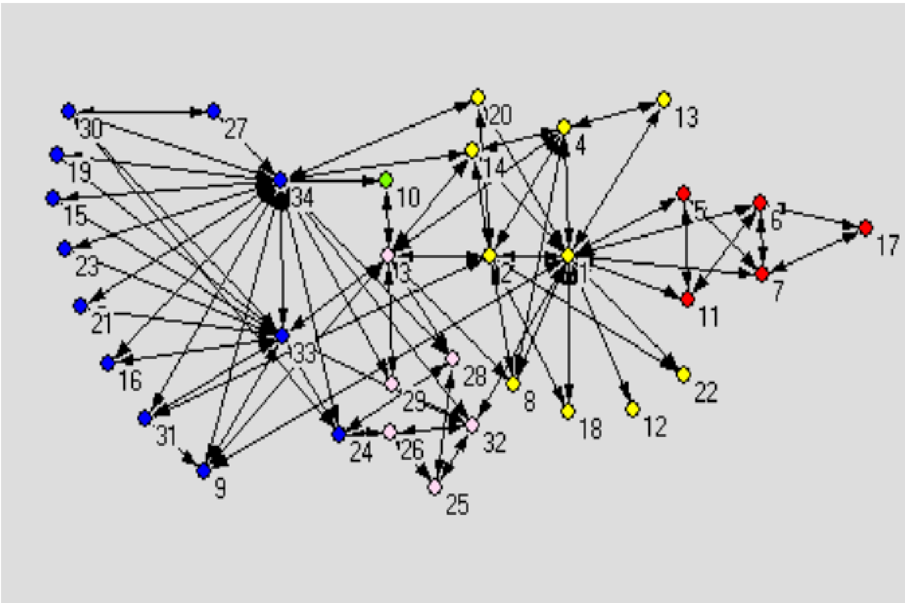

Figure 2: The Karate Club network and the optimal crisp cut given by Girvan and Newman.

Table 7: The Karate Club network. Comparative results in terms of the modularity measure $\widetilde{Q}$ and the number of groups.

Communities	GN	CMN	D&L	OUR FUZZY	OUR OCD	NFM	CFINDER
2	0.3599	0.3582	0.3148	0.2767	0.3392	**0.3714**	
3	0.3487	0.3717	0.38533	**0.34599**	0.3990	0.2498	
4	0.3632	**0.3879**	**0.4155**	0.3339	**0.4362**	0.2481	
5	**0.3850**	0.3800	0.4126	0.2922	0.4340	0.2707	
6	0.3517	0.3625	0.4063	0.2161	0.4049	0.1803	
7	0.3762	0.3547	0.3984	0.0472	0.3096	0.2501	
8	0.3583	0.3376	0.3885	0.0028	0.2151	0.2407	
9	0.3417	0.3201	0.3767	0.0026	0.2125	0.2407	
10	0.3247	0.3024	0.3624	0.0012	0.3176	0.2051	
11	0.3159	0.2843	0.3512	0.0006	0.2302	0.2550	
12	0.2986	0.2637	0.3344	0.0024	0.1196	0.2328	
13	0.2804	0.2543	0.3173	0.0001	0.2209	0.2748	
14	0.2628	0.2447	0.2999	0	0.2505	0.2223	
15	0.2475	0.2348	0.2822	0	0.3463	0.2715	0.3057
16	0.2268	0.2246	0.2515	0	0.2084	0.1362	
17	0.2089	0.2144	0.2308	0	0.1723	0.2715	
18	0.1898	0.2034	0.2109	0	0.1665	0.2413	
19	0.1812	0.1923	0.1846	0	0.2320	0.2728	
20	0.1600	0.1755	0.1741	0	0.2239	0.2021	
21	0.1469	0.1643	0.1637	0	0.1625	0.1694	
22	0.1203	0.1476	0.1443	0	0.1702	0.2715	
23	0.1081	0.1364	0.1116	0	0.1646	0.2944	
24	0.0090	0.1250	0.0854	0	0.1259	0.2021	
25	0.0080	0.1132	0.0762	0	0.3082	0.2021	
26	0.0069	0.0747	0.0659	0	0.2445	0.2715	
27	0	0.0563	0.0538	0	0.2128	0.1408	
28	0	0.0356	0.0301	0	0.2475	0.2021	
29	0	0.0265	0.0179	0	0.3048	0.2758	
30	0	0.0038	0.0079	0	0.1430	0.1674	
31	0	-0.008	-0.0027	0	0.1632	0.14	

Table 8: Fuzzy community detection with four classes and its crisp overlapping optimal solution with $\alpha = 0.3$.

Nodes	μ_{C_1}	μ_{C_2}	μ_{C_3}	μ_{C_4}	Crisp
1	0.1094	0.0894	0.0617	1	(0,0,0,1)
2	0.0187	0	0	1	(0,0,0,1)
3	0.1142	0	0.0509	1	(0,0,0,1)
4	0	0	0	1	(0,0,0,1)
5	0	1	0	0.0379	(0,1,0,0)
6	0	1	0	0.0016	(0,1,0,0)
7	0	1	0	0.0016	(0,1,0,0)
8	0	0	0	1	(0,0,0,1)
9	1	0	0	0.4282	**(1,0,0,1)**
10	1	0	0	0.6287	**(1,0,0,1)**
11	0	1	0	0.0379	**(0,1,0,1)**
12	0	0	0	1	(0,0,0,1)
13	0	0	0	1	(0,0,0,1)
14	0.0215	0	0	1	(0,0,0,1)
15	1	0	0	0	(1,0,0,0)
16	1	0	0	0	(1,0,0,0)
17	0	1	0	0	(0,1,0,0)
18	0	0	0	1	(0,0,0,1)
19	1	0	0	0	(1,0,0,0)
20	0.1251	0	0	1	(0,0,0,1)
21	1	0	0	0	(1,0,0,0)
22	0	0	0	1	(0,0,0,1)
23	1	0	0	0	(1,0,0,0)
24	0.7107	0	1	0	**(1,0,1,0)**
25	0	0	1	0	(0,0,1,0)
26	0	0	1	0	(0,0,1,0)
27	1	0	0	0	(1,0,0,0)
28	0.1136	0	1	0.1213	(0,0,1,0)
29	0.8937	0	0.7117	1	**(1,0,1,1)**
30	1	0	0.0255	0	(1,0,0,0)
31	1	0	0	0.2984	(1,0,0,0)
32	0.0931	0	1	0.1053	(0,0,1,0)
33	1	0	0.0007	0.0015	(1,0,0,0)
34	1	0	0.0167	0.0363	(1,0,0,0)

Table 9: Ranking (between brackets) of nodes according to their overlapping degree.

Nodes	2-overlap	3-overlap
1	0.1094 (9)	0.0894 (4)
2	0.0187 (16)	0 (8-34)
3	0.1142 (8)	0.0507 (5)
4	0 (20-34)	0 (8-34)
5	0.0379 (11)	0 (8-34)
6	0.0016 (17-18)	0 (8-34)
7	0.0016 (17-18)	0 (8-34)
8	0 (20-34)	0 (8-34)
9	0.4281 (4)	0 (8-34)
10	0.6286 (3)	0 (8-34)
11	0.0379 (12)	0 (8-34)
12	0 (20-34)	0 (8-34)
13	0 (20-34)	0 (8-34)
14	0.02152 (15)	0 (8-34)
15	0 (20-34)	0 (8-34)
16	0 (20-34)	0 (8-34)
17	0 (20-34)	0 (8-34)
18	0 (20-34)	0 (8-34)
19	0 (20-34)	0 (8-34)
20	0.1250 (6)	0 (8-34)
21	0 (20-34)	0 (8-34)
22	0 (20-34)	0 (8-34)
23	0 (20-34)	0 (8-34)
24	0.7149 (2)	0 (8-34)
25	0 (20-34)	0 (8-34)
26	0 (20-34)	0 (8-34)
27	0 (20-34)	0 (8-34)
28	0.1226 (7)	0.1149 (2)
29	0.8938 (1)	0.7081 (1)
30	0.0254 (14)	0 (8-34)
31	0.2984 (5)	0 (8-34)
32	0.1066 (10)	0.0942 (3)
33	0.0015 (19)	0.0007 (7)
34	0.0363 (13)	0.0166 (6)

6.2. Les Miserables network

Another classical network for testing algorithms can be found in *Les Miserables*. *Les Miserables* network is usually studied considering the graph associated to the relations among actors. In `http://www-personal.umich.edu/ mejn/netdata/`, a valuation of these relations can be found, that may be understood as the affinity or strength of such relations.

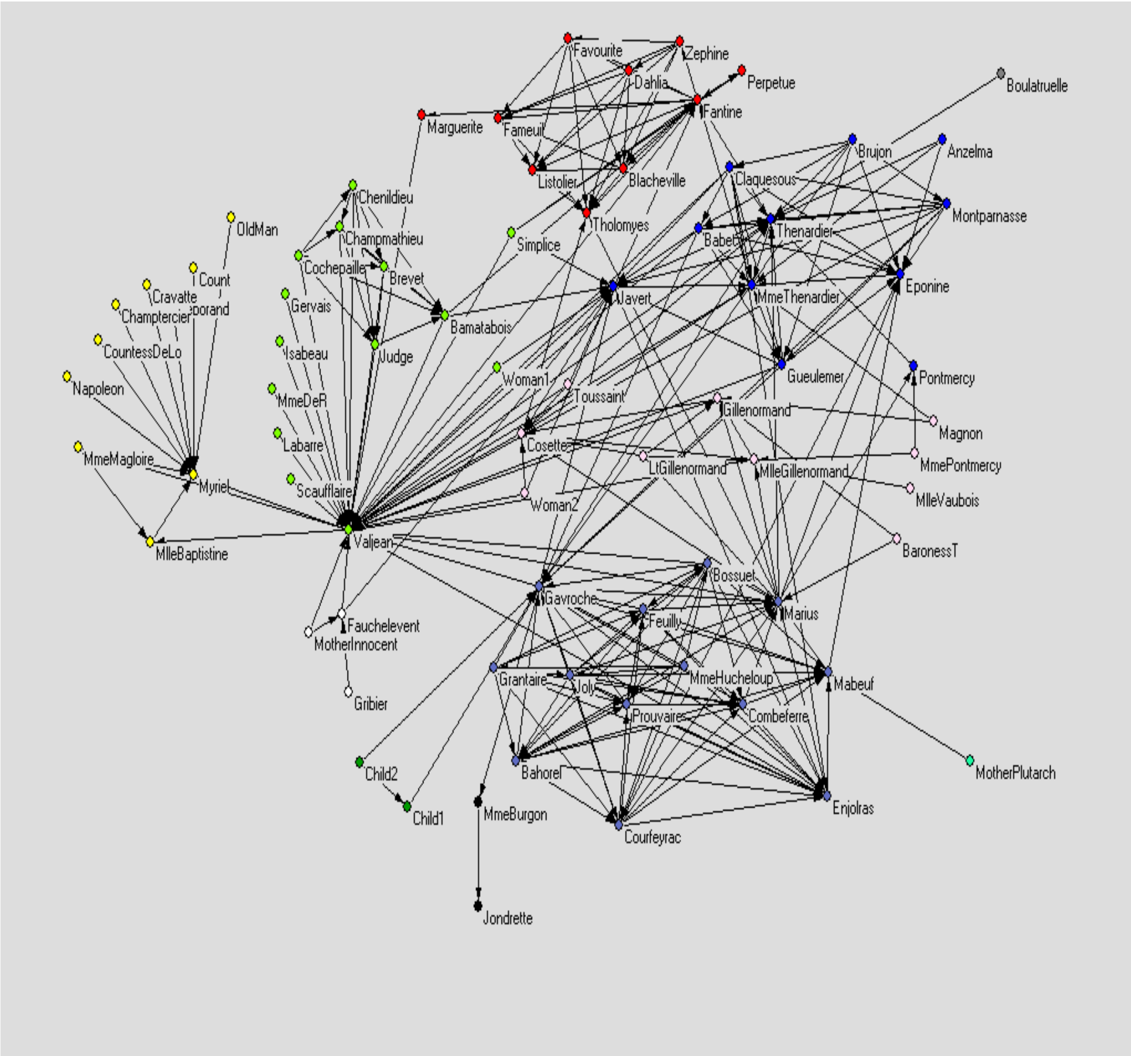

Figure 3: *LesMiserables* network and the optimal crisp cut given by Girvan and Newman.

Table 10 shows the performance of the chosen methods in *Les Miserables* network in terms of the proposed modularity measure $\widetilde{Q}$ for each number of communities. Again, the best result achieved by each method is shown in bold. Notice that the best results (among these algorithms) are obtained for the $OCD-z$ method proposed in this paper, that leads to a covering of the set of nodes in 7 communities achieving a modularity $\widetilde{Q} = Q_{GN} = 0.5641$. Particularly, again the proposed $OCD-z$ method obtains better results than the CCD algorithms, in which overlapping is not allowed. Let us remark that in this network there are some famous nodes (as for example *Cossete*) that clearly belong to more than one community. Thus, by allowing overlapping communities and nodes it is possible to improve in a significant way both the performance and the adequacy of the solution.

Notice that once more both $FCD-z$ and $OCD-z$ reach their optimum at different numbers of groups, 5 and 7 respectively. As in the previous experiment, we decided to present the $FCD-z$ clusters only for the number of groups leading to the best $OCD-z$ solution, that is for 7 groups. Table 11 shows the 7 fuzzy clusters provided by $FCD-z$, from which the corresponding optimal solution of $OCD-z$ can be easily obtained by taking the alpha-cut corresponding to $\alpha = 0.4$. The degrees of 2-overlap and 3-overlap derived from the previous clusters are shown in Table 12, again obtained with the squared root minimum and the maximum. Let us observe that, as in the previous $KarateClub$ example, the top-ranked nodes (i.e. those with the highest 2-overlap degree) in Table 12 coincide with the overlapping nodes detected by OND algorithms. Again, the results given by this rank are consistent with (and extend) the solutions given by classical algorithms as $CFINDER$, $CONGA$ or NFM. Also, notice that some nodes with a quite great 2-overlap degree (as node 45 $MotherInnocent$) however obtain a null or very low 3-overlap degree, while other nodes (as node 49 $Gavroche$) with a not-so-clear 2-overlap importance are however relatively important 3-overlapping nodes (notice also that, consistently with what we advanced in last paragraph, node 27 $Cosette$ presents the greatest degree of 3-overlapping). This shows the relevance of extending the notion of overlapping node to also consider potential overlaps of k (more than 2) communities.

Table 10: *Les Miserables* network. Comparative results in terms of the modularity measure $\widetilde{Q}$ for each number of groups.

Communities	**GN**	**CMN**	**D&L**	**OUR FUZZY**	**OUR OCD**	**NFM**	**CFINDER**
2	0.0746	0.3697	0.3718	0.2245	0.2331	0.0746	
3	0.2604	0.4421	0.4642	0.2353	0.2636	0.0819	
4	0.2660	**0.4569**	0.5110	0.3263	0.4939	0.2168	
5	0.4154	0.4468	0.5519	**0.4539**	0.5526	0.4740	
6	0.4587	0.4472	0.5542	0.4138	0.5564	0.4760	
7	0.4554	0.4545	**0.5561**	0.4218	**0.5641**	0.5081	
8	0.4536	0.4523	0.5556	0.1740	0.2764	**0.5214**	
9	0.4518	0.4501	0.5533	0.0019	0.2598	0.5186	
10	0.4524	0.4479	0.5508	0.0024	0.1130	0.5115	
11	**0.5380**	0.4457	0.5482	0.0008	0.2326	0.5168	
12	0.5347	0.4434	0.5455	0.0007	0.1741	0.5108	
13	0.5314	0.4412	0.5426	0.0004	0.1147	0.5071	
14	0.5281	0.4390	0.5396	0.0004	0.0607	0.5055	
15	0.5248	0.4367	0.5364	0.0001	0.0409	0.5032	
⋮							
44							0.4469
⋮							

Table 11: Fuzzy community detection with seven classes. The optimal crisp overlapping solution is reached with $\alpha = 0.4$.

Names	Nodes	μ_{C_1}	μ_{C_2}	μ_{C_3}	μ_{C_4}	μ_{C_5}	μ_{C_6}	μ_{C_7}
Myriel	1	1	0	0	0	0	0.0062	0
Napoleon	2	1	0	0	0	0	0	0
MlleBaptistine	3	1	0	0	0	0	0.6069	0
MmeMagloire	4	1	0	0	0	0	0.6069	0
CountessDeLo	5	1	0	0	0	0	0	0
Geborand	6	1	0	0	0	0	0	0
Champtercier	7	1	0	0	0	0	0	0
Cravatte	8	1	0	0	0	0	0	0
Count	9	1	0	0	0	0	0	0
OldMan	10	1	0	0	0	0	0	0
Labarre	11	0	0	0	0	0	0.875	0
Valjean	12	0.0022	0.0057	0.0073	0.01107	0.0043	0.875	0.0219
Marguerite	13	0	0.1743	0	0	0	0.875	0
MmeDeR	14	0	0	0	0	0	0.875	0
Isabeau	15	0	0	0	0	0	0.875	0
Gervais	16	0	0	0	0	0	0.875	0
Tholomyes	17	0	1	0	0.0354	0	0.0381	0
Listolier	18	0	1	0	0	0	0	0
Fameuil	19	0	1	0	0	0	0	0
Blacheville	20	0	1	0	0	0	0	0
Favourite	21	0	1	0	0	0	0	0
Dahlia	22	0	1	0	0	0	0	0
Zephine	23	0	1	0	0	0	0	0
Fantine	24	0	1	0.1381	0	0	0.4775	0.3569
MmeThenardier	25	0	0.02391	0	0	0.0175	0.1509	1
Thenardier	26	0	0.0179	0	0.0308	0.0131	0.1118	1
Cosette	27	0	0.2517	0	0.3560	0.6635	0.875	0.498
Javert	28	0	0.04020	0.0323	0.0689	0	0.5216	1
Fauchelevent	29	0	0	0	0	0	0.3964	1
Bamatabois	30	0	0.0285	1	0	0	0.0705	0.036
Perpetue	31	0	0.6247	0	0	0	0.875	0
Simplice	32	0	0.1681	0	0	0.875	0.5670	
Scaufflaire	33	0	0	0	0	0	0.875	0
Woman1	34	0	0	0	0	0	0.875	0.2862
Judge	35	0	0	1	0	0	0.0030	0
Champmathieu	36	0	0	1	0	0	0.0030	0
Brevet	37	0	0	1	0	0	0.0030	0
Chenildieu	38	0	0	1	0	0	0.0030	0
Cochepaille	39	0	0	1	0	0	0.0030	0
Pontmercy	40	0	0	0	0.0822	1	0	0.0830
Boulatruelle	41	0	0	0	0	0	0	1
Eponine	42	0	0	0	0.03097	0	0	1
Anzelma	43	0	0	0	0	0	0	1
Woman2	44	0	0	0	0	0	0.875	0.2355
MotherInnocent	45	0	0	0	0	0	0.875	0.9970
Gribier	46	0	0	0	0	0	0	1
Jondrette	47	0	0	0	0	0	0.125	0
MmeBurgon	48	0	0	0	0.9830	0	0.125	0
Gavroche	49	0	0	0	1	0	0.4609	0.42690
Gillenormand	50	0	0	0	0.0171	1	0.0219	0
Magnon	51	0	0	0	0	1	0	0.4201
MlleGillenormand	52	0	0	0	0.01117	1	0.0127	0
MmePontmercy	53	0	0	0	0	1	0	0
MlleVaubois	54	0	0	0	0	1	0	0
LtGillenormand	55	0	0	0	0.0014	1	0.0017	0
Marius	56	0	0.2744	0	1	0.431	0.7380	0.4445
BaronessT	57	0	0	0	0.0386	1	0	0
Mabeuf	58	0	0	0	1	0	0	0.0056
Enjolras	59	0	0	0	1	0	0.0293	0.0245
Combeferre	60	0	0	0	1	0	0	0
Prouvaire	61	0	0	0	1	0	0	0
Feuilly	62	0	0	0	1	0	0	0
Courfeyrac	63	0	0	0	1	0	0	0.00167
Bahorel	64	0	0	0	1	0	0	0
Bossuet	65	0	0	0	1	0	0.00291	0
Joly	66	0	0	0	1	0	0	0
Grantaire	67	0	0	0	1	0	0	0
MotherPlutarch	68	0	0	0	1	0	0	0
Gueulemer	69	0	0	0	0.0060	0	0.0156	1
Babet	70	0	0	0	0.006001214	0	0.0156	1
Claquesous	71	0	0	0	0.0057	0	0.0152	1
Montparnasse	72	0	0	0	0.0088	0	0.022	1
Toussaint	73	0	0	0	0	0	0.875	0.235
Child1	74	0	0	0	1	0	0	0
Child2	75	0	0	0	1	0	0	0
Brujon	76	0	0	0	0.03007	0	0	1
MmeHucheloup	77	0	0	0	1	0	0	0

Table 12: Ranking (between brackets) of nodes according to its overlapping degree.

Names	Nodes	2-overlap	3-overlap
Myriel	1	0.078 (34)	0 (21-77)
Napoleon	2	0 (44-77)	0 (21-77)
MlleBaptistine	3	0.7791 (5-6)	0 (21-77)
MmeMagloire	4	0.7791 (5-6)	0 (21-77)
CountessDeLo	5	0 (44-77)	0 (21-77)
Geborand	6	0 (44-77)	0 (21-77)
Champtercier	7	0 (44-77)	0 (21-77)
Cravatte	8	0 (44-77)	0 (21-77)
Count	9	0 (44-77)	0 (21-77)
OldMan	10	0 (44-77)	0 (21-77)
Labarre	11	0 (44-77)	0 (21-77)
Valjean	12	0.1480 (29)	0.1050(14)
Marguerite	13	0.4177 (16)	0 (21-77)
MmeDeR	14	0 (44-77)	0 (21-77)
Isabeau	15	0 (44-77)	0 (21-77)
Gervais	16	0 (44-77)	0 (21-77)
Tholomyes	17	0.1951 (23)	0.1883(9)
Listolier	18	0 (44-77)	0 (21-77)
Fameuil	19	0 (44-77)	0 (21-77)
Blacheville	20	0 (44-77)	0 (21-77)
Favourite	21	0 (44-77)	0 (21-77)
Dahlia	22	0 (44-77)	0 (21-77)
Zephine	23	0 (44-77)	0 (21-77)
Fantine	24	0.6909 (9)	0.5974 (4)
MmeThenardier	25	0.3881 (17)	0.1545(12)
Thenardier	26	0.3340 (19)	0.1754(10)
Cosette	27	0.8126 (3)	0.7059 (1)
Javert	28	0.7213 (8)	0.2624(7)
Fauchelevent	29	0.6293(12)	0 (21-77)
Bamatabois	30	0.2653 (21)	0.1907(8)
Perpetue	31	0.7906 (4)	0 (21-77)
Simplice	32	0.7535 (7)	0.4102 (5)
Scaufflaire	33	0 (44-77)	0 (21-77)
Woman1	34	0.5357 (13)	0 (21-77)
Judge	35	0.0556 (36)	0 (21-77)
Champmathieu	36	0.0556 (37-40)	0
Brevet	37	0.0556 (37-40)	0 (21-77)
Chenildieu	38	0.0556 (37-40)	0 (21-77)
Cochepaille	39	0.0556 (37-40)	0 (21-77)
Pontmercy	40	0.2894 (20)	0.2881(6)
Boulatruelle	41	0 (44-77)	0 (21-77)
Eponine	42	0.17602 (24)	0 (21-77)
Anzelma	43	0 (44-77)	0 (21-77)
Woman2	44	0.4855 (14)	0 (21-77)
MotherInnocent	45	0.9354 (1)	0 (21-77)
Gribier	46	0 (44-77)	0 (21-77)
Jondrette	47	0 (44-77)	0 (21-77)
MmeBurgon	48	0.3535 (18)	0 (21-77)
Gavroche	49	0.6788 (10)	0.65332 (3)
Gillenormand	50	0.1496 (27)	0.1322 (13)
Magnon	51	0.6498(11)	0 (21-77)
MlleGillenormand	52	0.1118 (33)	0.1047(15)
MmePontmercy	53	0 (44-77)	0 (21-77)
MlleVaubois	54	0 (44-77)	0 (21-77)
LtGillenormand	55	0.04216 (42)	0.0391 (20)
Marius	56	0.8591 (2)	0.6668 (2)
BaronessT	57	0.1977 (22)	0 (21-77)
Mabeuf	58	0.0751 (35)	0 (21-77)
Enjolras	59	0.1713 (26)	0.1566(11)
Combeferre	60	0 (44-77)	0 (21-77)
Prouvaire	61	0 (44-77)	0 (21-77)
Feuilly	62	0 (44-77)	0 (21-77)
Courfeyrac	63	0.0409 (43)	0 (21-77)
Bahorel	64	0 (44-77)	0 (21-77)
Bossuet	65	0.0539 (41)	0 (21-77)
Joly	66	0 (44-77)	0 (21-77)
Grantaire	67	0	0 (21-77)
MotherPlutarch	68	0	0 (21-77)
Gueulemer	69	0.1254 (30)	0.0776 (17-18)
Babet	70	0.1254 (31)	0.0776 (17-18)
Claquesous	71	0.1237 (32)	0.0762 (19)
Montparnasse	72	0.1494 (28)	0.0939 (16)
Toussaint	73	0.4855 (15)	0 (21-77)
Child1	74	0 (44-77)	0 (21-77)
Child2	75	0 (44-77)	0 (21-77)
Brujon	76	0.1735 (25)	0 (21-77)
MmeHucheloup	77	0 (44-77)	0 (21-77)

Remark 3. It is important to remark that the (crisp) overlapping solutions derived from the fuzzy ones have been obtained by choosing the best alpha-cut in terms of modularity (following the bi-level approach exposed in Section 5.2). In all the examples we have analyzed, we have always found an alpha-cut such that modularity increases from the FCD solution to the OCD one. A question to be explored is whether this is always the case, that is, whether always an alpha-cut exists or not that increases the modularity of the FCD solution. .

To conclude this paper, we would like to emphasize once again the importance of allowing the usage of overlap and grouping functions to define the proposed modularity measure, which naturally extends the classical Girvan-Newman modularity to the context of fuzzy communities without imposing a Ruspini partition of the set of nodes. As a consequence of this greater flexibility the methods proposed in this paper are able to deal with the following three problems at the same time and using the same modularity measure: the fuzzy community detection problem, the problem of ranking the overlapping nodes and the crisp identification of the overlapping nodes. Simultaneously addressing these three problems represents a clear advantage with respect to other approaches. Furthermore, the computational examples carried out on the classical $KarateClub$ and $LesMiserables$ networks shows that the proposed method obtains promising results. Last but not least, it is important to remark that the proposed modularity measure allows comparing partitions of different nature, as those provided by CCD, OCD and FCD methods.

Acknowledgment

This research was partially supported by the Government of Spain (grant TIN2012-32482), the Government of Madrid (grant S2013/ICCE-2845) and the UCM (Research Group 910149).

[1] A. Amo, J. Montero, E. Molina, Representation of consistent recursive rules. European Journal of Operational Research 130 (2001), 29-53.

[2] A. Amo, D. Gomez, J. Montero, G. Biging, Relevance and redundancy in fuzzy classification systems, Mathware and Soft Computing 8 (2001) 203-216.

[3] G. Beliakov, A. Pradera, T. Calvo, Aggregation Functions, a Guide to Practitioners, Springer-Verlag, Berlin, 2007.

[4] H. Bustince, J. Fernandez, R. Mesiar, J. Montero and R. Orduna, Overlap functions, Nonlinear Analysis-Theory Methods and Applications 72 (2010) 1488-1499.

[5] H. Bustince, M. Pagola, R. Mesiar, E. Hullermeier, F. Herrera, Grouping, Overlap, and Generalized Bientropic Functions for Fuzzy Modeling of Pairwise Comparisons, IEEE Transaction on Fuzzy Systems 20 (2012) 405-415.

[6] H. Bustince, J. Montero, E. Barrenechea, M. Pagola, Semiautoduality in a restricted family of aggregation operators, Fuzzy Sets and Systems 158 (2007) 1360-1377.

[7] T. Calvo, A. Kolesárová, M. Komorníková, R. Mesiar, Aggregation operators, properties, classes and construction methods. In T. Calvo et al. (Eds.): Aggregation Operators New trends ans Aplications, Physica-Verlag, Heidelberg, 2022, pp. 3-104.

[8] T. Calvo, G. Mayor, J. Torrens, J. Suñer, M. Mas and M. Carbonell, Generation of weighting triangles associated with aggregation fuctions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 8(4) (2000) 417-451.

[9] V. Cutello, J. Montero, Hierarchical aggregation of OWA operators: basic measures and related computational problems. Int. J. Uncertainty Fuzzinesss Knowledge based Systems 3 (1995), 17-26.

[10] V. Cutello, J. Montero. Recursive families of OWA operators. Proceedings FUZZ-IEEE Conference (IEEE Press, Piscataway, 1994), 1137-1141.

[11] V. Cutello, J. Montero. Recursive connective rules. International Journal of Intelligent Systems 14, 3-20, 1999.

[12] D. Gómez, J. Montero, A discussion of aggregation functions, Kybernetika 40 (2004) 107-120.

[13] Y. Dua, C. Gaoa, Y. Hub, S. Mahadevanc and Y. Deng, A new method of identifying influential nodes in complex networks based on TOPSIS. Physica A 399, 57-69 (2014).

[14] G. B. Davis and K. M. Carley, Clearing the FOG: fuzzy, overlapping groups for social networks. Social Networks 30 (2008) 201-216.

[15] S. Fortunato, Community detection in graphs. Physics Reports, 486(3) (2010) 75-174.

[16] M. Girvan and M.E.J. Newman, Community structure in social and biological networks Proc. Natl. Acad. Sci. USA 99 (2002) 7821.

[17] D. Gómez, J.R. Figueira, A. Eusebio, Modeling centrality measures in social network analysis using bi-criteria network flow optimization problems. European Journal of Operational Research 226 (2014) 354-365.

[18] D. Gómez, J. T. Rodríguez, J. Montero, H. Bustincec, E. Barreneche, n-Dimensional overlap functions. Fuzzy Sets and Systems 287 57-75.

[19] D. Gómez, K. Rojas, J.T. Rodríguez, J. Montero, G. Beliakov, Consistency and stability in aggregation operators: An application to missing data problems, International Journal of Computational Intelligence Systems 7 (2014) 595-604.

[20] D. Gómez, E. Zarrazola, J. Yáñez, J. Montero, A Divide-and-Link Algorithm for Hierarchical Clustering in Networks, Information Sciences, 316 (2015) 308-328.

[21] M. Grabisch, J. Marichal, R. Mesiar and E. Pap, Aggregation Functions. Encyclopedia of Mathematics and its Applications, Cambridge University, New York, 2009.

[22] S. Gregory, A fast algorithm to find overlapping communities in networks, Lecture Notes in Computer Science, 5211 408 (2008).

[23] S. Gregory, Fuzzy overlapping communities in networks, Journal of Statistical Mechanics: Theory and Experiment, P02017. (2011).

[24] A. Jurio, H. Bustince, M. Pagola, A. Pradera, R. R. Yager, Some properties of overlap and grouping functions and their application to image thresholding, Fuzzy Sets and Systems 229 (2013) 69-90.

[25] A. Kolesárová. Sequential aggregation. In: M. González et al. (Eds.), Proceedings of the Fifth International Summer School on Aggregation Operators, AGOP, Universitat de les Illes Balears, Palma de Mallorca, 2009, pp. 183-187.

[26] Z. Li, Y. Shi, and S. Chen, Exploring the Influence of Human Mobility on Information Spreading in Mobile Networks. International Journal of Moder Physics C. (2016) DOI: 10.1142/S0129183116500662.

[27] J. Liu, Fuzzy modularity and fuzzy community structure in networks. The European Physical Journal B-Condensed Matter and Complex Systems, 77(4), (2010) 547-557.

[28] J. Montero, D. Gómez, H. Bustince, On the relevance of some families of fuzzy sets. Fuzzy Sets and Systems. 158 (2007) 2429-2442.

[29] T. Nepusz, A. Petroczi, L. Negyessy and F. Bazso, Fuzzy communities and the concept of bridgeness in complex networks. Physical Review. E 77 016107. (2008)

[30] G. Palla, I. Derenyi, I. Farkas and T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society. Nature 435 (2004) 814.

[31] I. Psorakis, S. Roberts and B. Sheldon, Efficient Bayesian community detection using nonnegative matrix factorisation, arXiv:1009.2646 (2010)

[32] K. Rojas, D. Gómez, J.T. Rodríguez, J. Montero, Some properties of consistency in the families of aggregation functions. Advances in Intelligent and Soft Computing 107:169-176 (P. Melo-Pinto, P. Couto, C. Serodio, J. Fodor, B. De Baets, Eds.) Springer, Berlin (2011).

[33] K. Rojas, D. Gómez, J.T. Rodríguez, J. Montero, Strict Stability in Aggregation Operators. Fuzzy sets and Systems 228 (2013) 44-63.

[34] E.H. Ruspini, A new approach to clustering. Information and Control 15 (1969) 22-32.

[35] R.R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making. IEEE Transactions on Systems, Man and Cybernetics 18 (1988) 183-190.

[36] R.R. Yager and A. Rybalov, Nonconmutative self-identity aggregation. Fuzzy Sets and Systems 85 (1997) 73-82.

[37] R.R. Yager, Prioritized aggregation operators. International Journal of Approximate Reasoning 48 (2008) 263-274.

[38] J. Xie, S. Kelley, and B. K. Szymanski, Overlapping community detection in networks: The state-of-the-art and comparative study. ACM Computing Surveys (CSUR), 45(4) (2013) 43.

[39] W. Zachary. An information flow model for conflict and fission in small groups. Journal of Anthropological Research 33 (1977) 452-473.

[40] S. Zhang, R. Wang and X. Zhang, Identification of overlapping community structure in complex networks using fuzzy C-means clustering. Physica A 374 483 (2007).

[41] S. Zhang, R. Wang and X. Zhang, Uncovering fuzzy community structure in complex networks. Physical Review E (2007) 046103.