






## Article

# Community Detection Problem Based on Polarization Measures. An application to Twitter: the COVID-19 case in Spain

Inmaculada Gutiérrez<sup>1\*</sup> , Juan Antonio Guevara<sup>1</sup> , Daniel Gomez<sup>1,2</sup> , Javier Castro<sup>1,2</sup> , Rosa Espínola<sup>1,2</sup> 

<sup>1</sup> Faculty of Statistics, Complutense University Puerta de Hierro, s/n, Madrid, 28040, Spain

<sup>2</sup> Instituto de Evaluación Sanitaria, Complutense University, Madrid, Spain.

\* Correspondence: inmaguti@ucm.es

**Abstract:** In this paper we address one of the most important topics in the field of Social Networks Analysis: the community detection problem with additional information. That additional information is modeled by a fuzzy measure that represents the possibility of polarization. Particularly, we are interested in dealing with the problem of taking into account the Polarization of nodes in the community detection problem. Adding this type of information to the community detection problem makes it more realistic, as a community is more probably to be defined if the corresponding elements are willing to maintain a peaceful dialogue. The polarization capacity is modeled by a fuzzy measure based on the  $JDJ_{pol}$  measure of polarization related to two poles. We also present an efficient algorithm for finding groups whose elements are no polarized. Hereafter, we work in a real case. It is a network obtained from Twitter, concerning the political position against the Spanish government taken by several influential users. We analyze how the partitions obtained change when some additional information related to how polarized that society is, is added to the problem.

**Keywords:** Networks; Community Detection; Extended Fuzzy Graphs; Polarization; Fuzzy Sets; Ordinal Variation

## 1. Introduction

The field of Social Networks Analysis (SNA) encompasses a wide range of processes devoted to the investigation of social structures modeled by complex networks or graphs. These are models to show schemes of relations between the entities of a complex system, be it in technological applications, nature or society, so that the elements of the systems are described as vertices or nodes, and their interactions as links or edges. Particularly, online social networks are usually represented by a graph whose nodes are people and whose edges show relations of different nature: social, friendship, common interests, familiarities etc. Thousands of millions data are constantly generated, so the importance of the SNA has grown more and more in the recent decades, attracting the interests of many researchers from different areas. Generally, three analysis levels can be distinguished in SNA processes: the first, related to individuals; the next, related to the structures and relationships established by the graph structure, and the last, related to the analysis of interactions between previous levels.

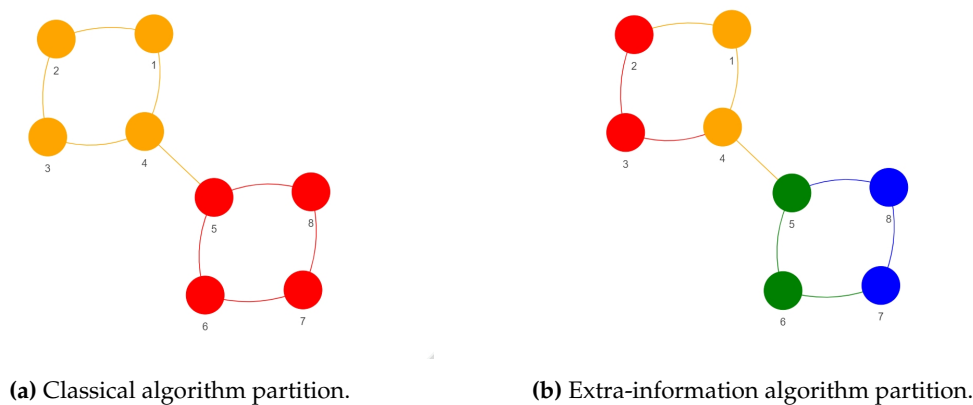
One of the features shown by complex networks is their internal group structure, a property which is far from being trivial. Trying to find these structures has become a highly relevant study topic in the SNA field: the well known community detection problem. This problem has evolved into an essential one, having many different applications in several areas. From the knowledge of the community structure of a complex network, several non-trivial internal features or organisations can be reached. Furthermore, it facilitates a better understanding of the dynamic processes which take place in the network and the inference of some properties or interactions between the elements.

Hence, the main goal of applying community detection algorithms to online social networks is to group individuals - represented by nodes - into communities, with the intention of knowing the

internal structure of a given society. In this light, community detection and social Polarization are closely related. In broader terms, Polarization can be understood as the split of a given population into two opposite groups, both with significant and similar size. Polarization measures [1–3] provide a single value which shows all these characteristics, taking into consideration some knowledge about the similarity between the individuals, the clusters in their population, etc. Thus, the structure of a given set of individuals impacts on the Polarization values shown by a given measure, as well as the presence of Polarization - or its degree - determines the topological structure of a network.

Because of the growing importance of the community detection problem, an extensive range of methods have been proposed to solve it [4–7], among which it is worth highlighting the Louvain [8] algorithm. It is a fast multi-phase method which provides good quality non-hierarchical partitions of the set of nodes, without a priori knowledge of the number of communities. Almost all the methods found in the literature, particularly the Louvain algorithm, have a point in common: the search of groups is based on the structural or topological characteristics of the graph. In this vein, the only information considered for the definition of groups is the knowledge represented by the graph, without deeming any additional data. Going a step further, several authors agree on the idea of adding additional information to the graph [9–12]. On our part, we consider the inclusion of some knowledge about the Polarization of the elements of a graph when grouping them. We agree on the importance of having groups whose elements are willing to peaceful dialogue, so that they are not prone to conflict.

Let us illustrate this idea with a basic example, consisting of a set of 8 elements connected as it is shown in Figure 1. Every community detection algorithm based on the graph structure, particularly on modularity optimization, organizes the elements into two clusters with 4 each one, by separating the two wheels, i.e.  $P = \{\{1, 2, 3, 4\}, \{5, 6, 7, 8\}\}$ . However, let us assume some knowledge about political position of the individuals against a government, represented by the vector  $O = (+, -, -, +, +, +, -, -)$ , so that if  $O_i = +$ , the individual  $i$  is in favour of the government, and the opposite happens when  $O_i = -$ . It is fair to accept that people who hold a similar political ideology, are less prone to conflict with each other than those who have opposite ideas. On this assumption, a desirable partition could be  $P^a = \{\{1, 4\}, \{2, 3\}, \{5, 6\}, \{7, 8\}\}$ .



**Figure 1.** Toy example: a graph with 8 nodes.

Concerning this idea of adding some additional information to the community detection problem, and based on methodology proposed by Gutierrez et al. [13–15] about the the community detection in graphs based on fuzzy measures, in this study we define a new method with the purpose of adding extra-information related to Polarization to the community detection problem in graphs. That additional information is based on the values given by the Polarization measure developed by Guevara et al. [3]. Our goal is to build a model consisting of network in combination with a polarization fuzzy measure whose structure fixes properly the reality. It is the polarization extended fuzzy graph, which takes both the attitude of the people and structure characteristics of the social network into consideration. On the basis of this model, we define a community detection method, which, by having

a graph and knowing the membership degree of every individual to two poles, provides realistic partitions of reality. The choice about several aggregation operators plays an essential role in this method, as it will be shown in the following pages.

The remainder of the paper is organized as follows. In Section 2 we set the basis of the paper, by introducing several concepts related to Graph Theory, fuzzy measures study and Polarization tools. Then, in Section 3 we work in the definition of a new fuzzy measure based on a Polarization measure. In combination with a graph, this fuzzy measure sets the definition of a new tool: the polarization extended fuzzy graph. In parallel, we define the non-polarization fuzzy measure, to represent the capacity of a set of elements to peacefully dialogue. From this non-polarization fuzzy measure and a crisp graph, we define the non-polarization extended fuzzy graph, for which we suggest a particular application in Section 4, related to searching partitions on it. We show the performance of this new methodology in a real case, working in the detection of groups in a polarization extended fuzzy graph whose origin is Twitter. The experiment design and the methodology can be found in Section 5. We finish this paper in Sections 6 and 7, showing some discussion and conclusions about the work done.

## 2. Preliminaries

In this section we introduce several concepts on which this paper is based. We divide it into two main parts: one is related to networks and graphs as well as the community detection problem, and the other is related to Polarization background.

### 2.1. Graphs, fuzzy graphs and extended fuzzy graphs

Let  $V = \{1, \dots, n\}$  denote a finite set of elements, which will be named as nodes or vertices. Let  $E = \{\{i, j\} \mid i, j \in V\}$  denote a finite set of edges which represents the set of related elements of  $V$ , so that  $\exists e = \{i, j\} \in E$  if and only if the nodes  $i, j \in V$  are related or connected. Then, the pair  $G = (V, E)$  is a graph or network (both terms will be used interchangeably throughout the paper).

A graph is usually referred to as the pair  $G = (V, E)$ ; however, there is another option for its representation: its adjacency matrix, usually denoted as  $A$ . This matrix represents the direct connections among all the pairs of elements of  $V$  so that the element  $A_{ij} \neq 0$  if and only if  $\exists e = \{i, j\} \in E$ . Moreover, it can be assumed the existence of a non-negative weights function defined over the set of edges  $E$ . In this case, the element  $A_{ij}$  represents the weight of the edge  $e = \{i, j\}$ . These graphs with a weights function defined over the set of edges are known as weighted graphs.

Beyond this classical concept of graph, Rosenfeld introduced the fuzzy graphs [16], based on the fuzzy relations among the individuals. This new notion was settled on the basis of the fuzzy sets, firstly introduced by Zadeh [17]. A fuzzy set is somewhat like a set but defined as a generalized extension of the classical term. So the elements of a fuzzy set have a membership degree to it.

**Definition 1** (Fuzzy set [17]). A fuzzy set  $\tilde{A}$  over the domain  $X$  is defined as

$$\tilde{A} = \{(x, \eta_A(x)) \mid x \in X\}$$

where  $\eta_A$  denotes a membership degree function, i.e.  $\eta_A : X \longrightarrow [0, 1]$ .

Far from the classical notion of crisp belonging, the membership degree of the elements related to fuzzy sets facilitates the consideration of the relations between the individuals intended by the definition of the fuzzy graphs. This tool, very useful to model situations in which there is some vagueness or uncertainty about the representation of the knowledge, has been widely used in many fields [18–20]. The consideration of several aggregation functions, particularly the conjunction operators [21] plays an essential role in the definition of the fuzzy graphs, as it is shown below.

**Definition 2** (Fuzzy graph [16]). *Let us consider the set  $V$  and the functions  $\eta : V \rightarrow [0, 1]$  and  $\psi : V \times V \rightarrow [0, 1]$  such that for every  $x, y \in V$  the property  $\psi(x, y) \leq B(\eta(x), \eta(y))$  holds, being  $B$  a conjunction operator. A fuzzy graph is a triplet  $\hat{G} = (V, \eta, \psi)$ , where  $\eta$  is called fuzzy vertex set, and  $\psi$  is called fuzzy edge set of  $\hat{G}$ .*

Note that if  $\eta(x) = 1$  for every  $x \in V$  previous constraints trivially hold since, as  $B$  is an aggregation (conjunction) operator, then  $B(\eta(x), \eta(y)) = B(1, 1) = 1$ . Then, because of the definition of  $\psi$ , it trivially holds that  $\psi(x, y) \leq B(\eta(x), \eta(y)) = B(1, 1) = 1$ . On this basis, previous definition can be simplified, by assuming that the fuzzy vertex set is crisp, so that the whole information is provided by the pair  $\hat{G} = (V, \psi)$ .

**Definition 3** (Crisp graph with fuzzy edges [22]). *Let us consider the graph  $G = (V, E)$  and the function  $\psi : E \rightarrow [0, 1]$ , which is a fuzzy set defined in the set of links  $E$ . The triplet  $\hat{G} = (V, E, \psi)$  is a fuzzy graph or a crisp graph  $G$  with fuzzy edges  $\psi$ .*

**Remark 1.** *Let us emphasize that, with a degree function  $\psi$ , we can easily obtain the set  $E = \{(x, y) \in V \times V \mid \psi(x, y) > 0\}$ , so, a crisp graph with fuzzy edges could be understood as the set of fuzzy graphs with the crisp nodes  $V$ . At this point, a fuzzy edge has to be null if the corresponding link does not exist in  $E$ . Under this assumption, both definitions of fuzzy graph are equivalent.*

Note that, considering this characterization, only nodes which are adjacent in the crisp graph  $G$  can have any membership degree in  $\psi$ . For this reason, from a mathematical point of view, this type of graph could be understood as a kind of weighted graph in which the weights represent the membership degree of the edges. An amplified vision of this model was introduced in [13] by means of the extended fuzzy graph, a concept based on fuzzy measures. Fuzzy measures are a type of monotonic set functions really useful to approach the reality. These functions allow us to model and represent the capacity not only of pairs of elements, but also of set with any cardinal. The use of these functions notably generalizes the modeling of knowledge, with the consideration of relations and synergies which can not be represented with other tools. Moreover, fuzzy measures are really sensible to the management of vagueness and uncertainty, so these are a good option to face the modeling of relations between elements. As it is pointed in [23], fuzzy/capacity measures are fundamental in modelling dependencies among the inputs and constitute a natural tool for modelling in multiple criteria decision analysis, aggregation, group decision making or game theory.

**Definition 4** (Fuzzy Measure [24]). *Let  $V$  denote a non empty set. A fuzzy measure is a set function  $\mu : 2^V \rightarrow [0, 1]$  for which the following holds.*

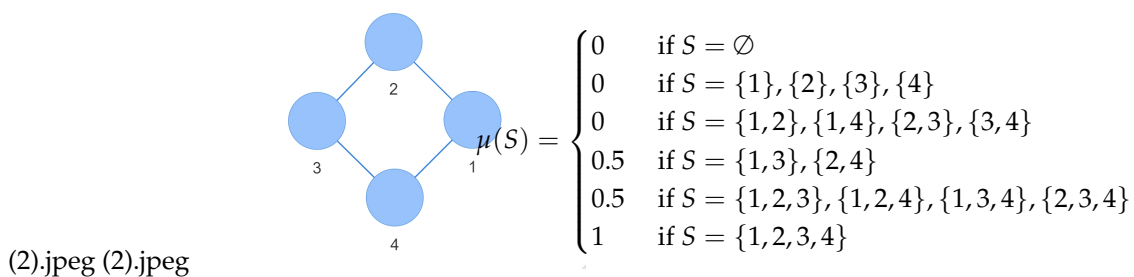
- $\mu(\emptyset) = 0$
- $\mu(V) = 1$
- $\mu(A) \leq \mu(B)$ ,  $\forall A, B \subseteq V$  such that  $A \subseteq B$

Then, with the combination of the ability of the graph to model connections between elements, and the ability of the fuzzy measures to handle the capacity related to any set of elements, it was defined the extended fuzzy graph. This tool is a graph together with a fuzzy measure defined over the set of nodes. The incorporation of a fuzzy measure goes far from the previous notion of fuzzy graphs, which are limited to the consideration of pairs of elements. In this vein, by means of a fuzzy measure defined over the set of nodes, we can represent situations in which more than two nodes are implied, independent of the way they are connected through the graph. It is obvious that the representation ability of the extended fuzzy graph goes far from that of the existing tools, so that much more complex situations can be addressed, with a proper modeling of reality.

**Definition 5** (Extended fuzzy graph [13]). Let  $G = (V, E)$  denote a graph, and let  $\mu : 2^V \rightarrow [0, 1]$  denote a fuzzy measure defined over the set of nodes  $V$ . An extended fuzzy graph is a triplet  $\tilde{G} = (V, E, \mu)$ , also called crisp graph with fuzzy measure  $\mu$ .

In the following example we show how it is possible to represent complex situations with several information sources by means of an extended fuzzy graph.

**Example 1.** Let us consider the graph  $G = (V, E)$  with 4 nodes (Figure 2). We assume some knowledge about the political position of the individuals against a government, represented by the vector  $O = (+, -, +, -)$ , so that if  $O_i = '+'$ , the individual  $i$  is in favour of the government, and the opposite happens if  $O_i = '-'$ . These are strong political opinions, so it is not easy for individuals with opposite ideas to peaceful dialogue. However, when two individuals with the same idea are together, they can discuss peacefully at great length. Let the fuzzy measure  $\mu : 2^V \leftarrow [0, 1]$  represent somehow the capacity of each feasible group of elements to discuss depending on their ideology. With the extended fuzzy graph  $\tilde{G} = (V, E, \mu)$  we represent the connections between the individuals as well as their ability to peaceful dialogue regarding their political ideas.



**Figure 2.** Extended fuzzy graph  $\tilde{G} = (V, E, \mu)$ .

## 2.2. Community detection problem

One of the main applications of graphs is related to the community detection problem. Many complex networks usually have an intern modular structure so that the nodes are organized into modules with dense internal connections, scarcely interconnected externally. The goal of community detection problem is to find these hidden structures, i.e., to establish a *good* partition of the set of nodes.

Some authors understand the community detection problem as an optimization problem [4,25]. The modularity  $Q$  is one of the most used measures as objective function to be optimized. This measure, whose value is determined by the topology of the network, is used to quantify the goodness of a partition. It was firstly defined by Newman and Girvan [26] as follows.

**Definition 6** (Modularity  $Q$  [26]). Let  $G = (V, E)$  denote a graph, with  $m = |E|$  edges, and let  $P$  denote a partition of the set of nodes  $V$ . The modularity of  $P$  is defined as:

$$Q(P) = \frac{1}{2m} \sum_{i,j \in V} \left[ A(i, j) - \frac{k_i k_j}{2m} \right] \delta(C_i, C_j) \quad (1)$$

where  $A$  denotes the adjacency matrix of  $G$ ;  $k_i$  denotes node  $i$ 's degree, and  $C_i$  denotes the cluster to which node  $i$  belongs. The value of  $\delta(C_i, C_j)$  is 1 if nodes  $i$  and  $j$  are in the same cluster ( $C_i = C_j$ ), and 0 otherwise.

Many approaches have been proposed in the last decades to face the community detection problem [4–7,25]. It is worth highlighting the Louvain algorithm [8], one of the most popular methods in this field, proposed by Blondel et al. in 2008. This algorithm performs very well, particularly with large networks, for which good quality non-hierarchical partitions are detected in a very little computing time. It is an iterative multi-phase method, based on modularity optimization and local moving [27], for which the variation of modularity,  $\Delta Q_i(j)$  defined in [8], is a key element. This

variation represents the gain attained if the node  $i$  is moved to the community to which  $j$  belongs, and it is calculated in each step of the Louvain algorithm until a maximum of modularity is reached. The Louvain algorithm is a key point of the methodology which will be proposed in the following pages, related to the community detection in graphs with additional information.

**Example 2.** Let us recall the graph of 8 nodes presented in the introduction (Figure 1). There we affirmed that if the aim is to maximize the modularity, the partition should be  $P = \{\{1, 2, 3, 4\}, \{5, 6, 7, 8\}\}$  (particularly, this partition is obtained with the Louvain algorithm). Observe that, indeed,  $Q(P) = 0.3889 > 0.1914 = Q(P^a)$ , where  $P^a = \{\{1, 4\}, \{2, 3\}, \{5, 6\}, \{7, 8\}\}$  is a desirable partition that could be obtained if the additional information defined by that vector  $O$  were considered.

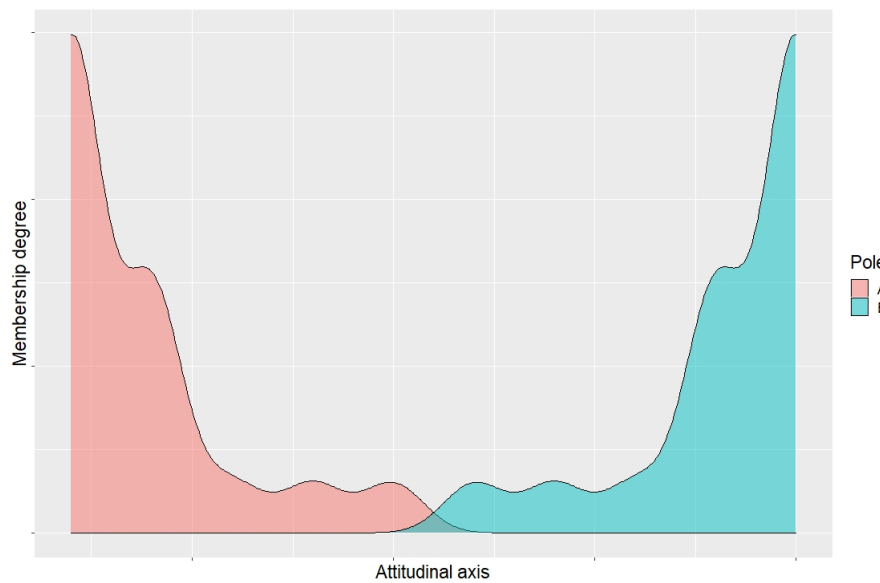
### 2.3. Polarization

In the last few decades, both the concept of Polarization and its measurement have aroused an increasingly interested in the literature. Due to the new digital technologies, the Web 2.0 and the social big data analysis, the study of the social conflict is now more reachable than ever. In broader terms, Polarization can be understood as the split of a given society into two different and opposite groups along an attitudinal axis. The measurement of the Polarization is studied in several disciplines [1,28–30]. Nevertheless, we find a lack of consensus on the literature about how Polarization should be measured. Different works related to political Polarization, income Polarization or social Polarization, differ from each other on which sub-concepts belong to the Polarization phenomenon. In the framework of social sciences, the measurement of the Polarization formally started by the end of the twentieth century with [1,31]. Since then, a growing number of diverse Polarization measures has arisen. Some of them, based on [1,31], are related to the concept of inequality; others rooted in different approaches, such as those based on the distance measurement of the bimodality, related to measuring ethnic and religious diversity [32]; or those related to the modeling of the public opinion considering the individual heterogeneity [33]. In Guevara et al. [3] the authors introduced a Polarization measure based on the fuzzy set approach, with which it is possible to avoid the duality Yes/No. Due to the fuzzy sets nature, this measure can deal with numeric, ordinal or linguistic variables as well. The main argumentation of that work is based on the assumption that ‘reality is not black and white’. When considering the classical Polarization measures found in the literature, each individual is forced to belong to a specific position along the Polarization axis [1,2]. The polarization measure introduced in [3] is aimed to overcome these issues. In the paragraphs below we present the ideas on which that paper is based; then we show the definition of the measure itself.

The way in which the information is aggregated is one of the main basis of this measure. In [3], Aggregation Operators (AO) [17] are used for this aim. AO were originally defined to aggregate the resulting values of the membership functions of a fuzzy set. Particularly, overlap functions [21] are used in this measure to show the degree  $z$  of the intersection of both classes with respect to the object  $c$ . On the opposite, grouping functions [34] are used to get the degree  $z$  up to which the combination of these classes is supported. Let us detail the characterization of the  $JDJ_{pol}$  measure.

We consider the finite set  $V$  and the one-dimensional variable  $X$  (ordinal or numeric). We assume that  $X$  has two extreme and opposite values or poles,  $X_A$  and  $X_B$ . Then, regarding the value of each element of  $V$  on  $X$ , we can measure their membership degree to each of these the poles  $X_A$  and  $X_B$ .

These membership degrees are represented by the membership functions  $\eta_{X_A}, \eta_{X_B} : V \rightarrow [0, 1]$ , so that, for every  $i \in V$ ,  $\eta_{X_A}(i)$  and  $\eta_{X_B}(i)$  represent the membership degree of the element  $i$  to the extreme pole  $X_A$  and to the extreme pole  $X_B$ , respectively.



**Figure 3.** Example of Bi-Polarization

In this scenario, Polarization exists when almost half the population is placed by the extreme position  $X_A$ , and the other half is placed by the extreme position  $X_B$ . In the Figure 3 we show an illustrative example of two membership degree functions  $\eta_{X_A}$  and  $\eta_{X_B}$ .

In this context, in [3] the  $JDJ_{pol}$  polarization measure was defined as the expected risk of polarization of a given population. In this vein,  $JDJ_{pol}$  measures the risk of polarization for each pair of individuals  $\{i, j\}$  of a finite set  $V$ . The obtained value is given by the summation of all those comparisons between all pairs and its aggregation. This value depends on the following:

- The closeness of the element  $i$  to the pole  $X_A$  and the closeness of the node  $j$  to the pole  $X_B$ , represented by  $\eta_{X_A}(i)$  and  $\eta_{X_B}(j)$  respectively.
- The closeness of the element  $j$  to the pole  $X_A$  and the closeness of the node  $i$  to the pole  $X_B$ , represented by  $\eta_{X_A}(j)$  and  $\eta_{X_B}(i)$  respectively.
- The grouping operator  $\varphi$  chosen.
- The overlapping operator  $\phi$  chosen.

Let us provide a mathematical definition of the  $JDJ_{pol}$  measure.

**Definition 7** ( $JDJ_{pol}$  Polarization measure [3]). Let  $V$  denote a finite set, and let  $\eta_{X_A}$  and  $\eta_{X_B}$  denote the membership functions of the elements of  $V$  to the extreme poles  $X_A$  and  $X_B$ . Let  $\varphi : [0, 1]^2 \rightarrow [0, 1]$  denote a grouping operator and  $\phi : [0, 1]^2 \rightarrow [0, 1]$  denote an overlapping operator. Then,  $JDJ_{pol}$  measure is defined as:

$$JDJ_{pol}(V, \eta_{X_A}, \eta_{X_B}, \varphi, \phi) = \sum_{\substack{i, j \in V \\ i \leq j}} \varphi(\phi(\eta_{X_A}(i), \eta_{X_B}(j)), \phi(\eta_{X_B}(i), \eta_{X_A}(j))) \quad (2)$$

**Remark 2.** The membership degrees defined by the a membership function are always non-negative, particularly those degrees concerning  $\eta_{X_A}$  and  $\eta_{X_B}$ . Then, because of the properties of the grouping and overlapping operators, the measure  $JDJ_{pol}$  is monotone and non-negative.

The performance of  $JDJ_{pol}$  shows the highest values of Polarization in those cases in which the 50% of the elements are located in one extreme value of the attitudinal axis and the other 50% of the elements are located in the opposite extreme value.

**Example 3.** For a better understanding of the polarization measure proposed in [3], let us show an example in which the opinion of individuals are crisp and opposite. Let  $V$  denote a set of individuals

$V = \{1, 2, 3, 4, 5, 6, 7, 8\}$ , and let us assume that we know their membership degree towards a given attitudinal axis  $X$  expressed by means of two membership functions as  $(\eta_{X_A}(1), \dots, \eta_{X_A}(8)) = (1, 0, 0, 1, 1, 1, 0, 0)$  and  $(\eta_{X_B}(1), \dots, \eta_{X_B}(8)) = (0, 1, 1, 0, 0, 0, 1, 1)$ , related to the poles  $X_A$  and  $X_B$  respectively. In this case, half the population have a strong feeling of identification towards the pole  $X_A$  while the other half feels identified towards the pole  $X_B$ . We choose the maximum as the grouping operator  $\varphi$ , and the product as the overlapping operator  $\phi$ . Then, two different situations can be distinguished:

1. **Comparing nodes with different membership degrees:**  $\eta_{X_A}(i) \neq \eta_{X_A}(j)$  and  $\eta_{X_B}(i) \neq \eta_{X_B}(j)$ .

Since this is the case for nodes 1 and 2, let us compute the  $JDJ_{pol}$  value for them. We have that  $\eta_{X_A, X_B}(1) = (\eta_{X_A}(1), \eta_{X_B}(1)) = (1, 0)$  and  $\eta_{X_A, X_B}(2) = (\eta_{X_A}(2), \eta_{X_B}(2)) = (0, 1)$ . Then we have to aggregate by a grouping function ( $\varphi$ ) two values: the degree to which the element 1 belongs to  $X_A$  and the element 2 belongs to  $X_B$ ,  $a = \phi(\eta_{X_A}(1), \eta_{X_B}(2)) = \phi(1, 1) = 1$ ; and the degree to which the element 1 belongs to  $X_B$  and the element 2 belongs to  $X_A$ ,  $b = \phi(\eta_{X_B}(1), \eta_{X_A}(2)) = \phi(0, 0) = 0$ . Finally, we verify which one of these two facts are true by a grouping function  $\phi(1, 1) = 1$ , since this is a case with high polarization.

2. **Comparing two nodes with the same membership degrees:**  $\eta_{X_A}(i) = \eta_{X_A}(j)$  and  $\eta_{X_B}(i) = \eta_{X_B}(j)$ .

This is the case of elements 5 and 6, for which we compute the  $JDJ_{pol}$ . We have that  $\eta_{X_A, X_B}(5) = (\eta_{X_A}(5), \eta_{X_B}(5)) = (1, 0)$  and  $\eta_{X_A, X_B}(6) = (\eta_{X_A}(6), \eta_{X_B}(6)) = (1, 0)$ . Then we have to aggregate by a grouping function ( $\varphi$ ) two values: the degree to which the element 5 belongs to  $X_A$  and the element 6 belongs to  $X_B$ ,  $a = \phi(\eta_{X_A}(5), \eta_{X_B}(6)) = \phi(1, 0) = 0$ ; and the degree to which the element 5 belongs to  $X_B$  and the element 6 belongs to  $X_A$ ,  $b = \phi(\eta_{X_B}(5), \eta_{X_A}(6)) = \phi(0, 1) = 0$ .

Following the method explained above, we compute the final value of the measure  $JDJ_{pol}$  concerning the set  $V$ , making all the comparisons between each node of the set. As the length of the set is equal to 8, a total of 64 comparison will be needed. Note we can dispense with self comparisons and double comparisons between nodes, so finally those 64 comparisons are reduced to 28.

In this example, each comparison between nodes with same membership degrees will result in 0 value. In contrast, the comparison between nodes with opposite membership degrees will result in 1. So that, the set is split into two groups; the final result is the sum of sixteen 1s and sixteen 0s. Thus, in this case  $JDJ_{pol}(V, \eta_{X_A}, \eta_{X_B}, \max, \text{product}) = 16$ . It is worth highlighting that in this case,  $JDJ_{pol}$  is equal to the variance of a binomial variable. In addition, for a better interpretation of this result, the measure proposed in [3] could be normalized. For the normalizing process, only those comparisons which show values higher than 0 are considered due that it does not make any sense adding a null value (i.e.  $JDJ_{pol}^N = \frac{16}{16} = 1$ , resulting the highest value for a polarization index).

As it can be seen in the last example, despite that the traditional way to measure a Polarization value is to have the values of a given feature, using the Polarization measure proposed in [3] allows us to compute the Polarization value only knowing the membership degree. Due to this characteristic, we can measure the Polarization with different types of variables (e.g.: numeric, ordinal, categorical, etc.).

### 3. Networks with additional information: the polarization extended fuzzy graph

In this section we show that the Polarization measure defined in [3] can be seen as a fuzzy measure when it is computed for each subset of individuals. We work in the idea of adding some additional information to a given graph. To carry on with it, we assume the existence of a crisp graph  $G = (V, E)$  and some knowledge about the attitude of the elements of  $V$  concerning any particular issue. Firstly we define a fuzzy measure emanated from the Polarization measure relative to this attitudinal knowledge. This fuzzy measure reveals the fuzzy relations existing between all the pairs of elements of  $V$  emanated from the measure  $JDJ_{pol}$  [3].

**Definition 8** (Polarization fuzzy measure  $\mu_P$ ). Given a uni-dimensional variable, let  $V$  denote a set of  $n$  individuals, about which we know the membership degree to the extreme poles of that variable,  $X_A$  and  $X_B$ , represented by the membership functions  $\eta_{X_A}$  and  $\eta_{X_B}$  respectively. Let the functions  $\phi : [0, 1]^2 \rightarrow [0, 1]$  and

$\varphi : [0, 1]^2 \rightarrow [0, 1]$  denote a grouping operator [34] and an overlapping operator [21], respectively. Let  $S$  denote a subset of  $V$ , and let  $JDJ_{pol}(\{i, j\}, \eta_{X_A}, \eta_{X_B}, \varphi, \phi) = \varphi(\phi(\eta_{X_A}(i), \eta_{X_B}(j)), \phi(\eta_{X_A}(j), \eta_{X_B}(i)))$ , according to the equation 2. We define the polarization fuzzy measure  $\mu_{P-}$  as follows:

$$\mu_{P-}(S) = \frac{JDJ_{pol}(S, \eta_{X_A}, \eta_{X_B}, \varphi, \phi)}{JDJ_{pol}(V, \eta_{X_A}, \eta_{X_B}, \varphi, \phi)} \quad (3)$$

**Proposition 1.** The function  $\mu_{P-}$  characterized in the Definition 8 is a fuzzy measure

**Proof.** To demonstrate this affirmation, we will show that the properties enunciated in the Definition 4 concerning fuzzy measures hold for  $\mu_{P-}$ .

- $\mu_{P-}(\emptyset) = 0$ . Trivial.
- $\mu_{P-}(V) = 1$ .  $\mu_{P-}$  is 1-normalized by definition.
- Let  $A, B \subseteq V$  such that  $A \subseteq B$ . Then,  $\mu_{P-}(A) \leq \mu_{P-}(B)$ . By definition,  $JDJ_{pol}$  is a monotonic measure, so this property trivially holds.

□

**Remark 3.** Note that previous definition of the polarization fuzzy measure  $\mu_{P-}$  could be re-formulated as a summation concerning the different pairs of elements, i.e.

$$\mu_{P-}(S) = \sum_{i,j \in S} P_{i,j}^- \quad (4)$$

where

$$P_{i,j}^- = \frac{\varphi(\phi(\eta_{X_A}(i), \eta_{X_B}(j)), \phi(\eta_{X_B}(i), \eta_{X_A}(j)))}{JDJ_{pol}(V, \eta_{X_A}, \eta_{X_B}, \varphi, \phi)} \quad (5)$$

Because of the properties of  $\mu_{P-}$ ,  $P^-$  is symmetric, non-negative, normalized, and its main diagonal is null.

Because of the interpretation of the measure  $JDJ_{pol}$ ,  $P_{i,j}^-$  represents the risk of conflict concerning the elements  $i$  and  $j$ . So that,  $\mu_{P-}$  represents the capacity of the elements to argue, to trigger conflict and arguments. Hence, it is a recommended model to properly represent the discrepancy or distance between the individuals.

**Example 4.** In this example we show the calculation of  $\mu_{P-}$  for a given set  $V$  with 4 elements. We consider the membership functions  $\eta_{X_A}$  and  $\eta_{X_B}$  defined in Table 4. We consider the functions  $\varphi = \max$  and  $\phi = \text{product}$ .

**Table 1.** Membership degree of each element of  $V$  to the poles  $X_A$  and  $X_B$ .

Element	$\eta_{X_A}$	$\eta_{X_B}$
1	1	0
2	0	1
3	1	0
4	0	1

Note that  $JDJ_{pol}(V) = 4$  is the amount of arguments among the 4 elements, i.e. the capacity to trigger conflict. These conflicts come from the groups  $\{1, 2\}$ ,  $\{1, 4\}$ ,  $\{2, 3\}$  and  $\{3, 4\}$ .

**Table 2.** Values of the fuzzy measures  $\mu_{P-}$ .

S	$\{1, 2\}$	$\{1, 3\}$	$\{1, 4\}$	$\{2, 3\}$	$\{2, 4\}$	$\{3, 4\}$	$\{1, 2, 3\}$	$\{1, 2, 4\}$	$\{1, 3, 4\}$	$\{2, 3, 4\}$	$\{1, 2, 3, 4\}$
$\mu_{P-}(S)$	0.25	0	0.25	0.25	0	0.25	0.5	0.5	0.5	0.5	1

**Remark 4.** We can define a measure obtained from the negation of the risk of polarization between two elements. That measure will have an opposite meaning than the capacity obtained from the  $JDJ_{pol}$ . Let  $N : [0, 1] \rightarrow [0, 1]$  denote a negation aggregator, and let us define

$$\widetilde{JDJ}(\{i, j\}, \eta_{X_A}, \eta_{X_B}, \varphi, \phi) = N(\varphi(\phi(\eta_{X_A}(i), \eta_{X_B}(j)), \phi(\eta_{X_A}(j), \eta_{X_B}(i)))) \quad (6)$$

Then, we define the matrix  $P^+$  as

$$P_{ij}^+ = \frac{\widetilde{JDJ}(\{i, j\}, \eta_{X_A}, \eta_{X_B}, \varphi, \phi)}{\sum_{r, s \in V} \widetilde{JDJ}(\{r, s\}, \eta_{X_A}, \eta_{X_B}, \varphi, \phi)}, \quad i, j \in V \quad (7)$$

**Definition 9** (Non-polarization fuzzy measure  $\mu_{P^+}$ ). Given a finite set  $V$ , a grouping function  $\varphi$ , a conjunction function  $\phi$ , a negation operator  $N$  and two membership functions  $\eta_{X_A}, \eta_{X_B} : V \rightarrow [0, 1]$ , let  $P^+$  be the matrix characterized in equation (7). Then, from matrix  $P^+$ , we can define a measure which represents the capacity of the elements of a set to peacefully dialogue without risk of Polarization:

$$\mu_{P^+}(S) = \sum_{i, j \in S} P_{ij}^+ \quad (8)$$

**Remark 5.** Trivially,  $\mu_{P^+}$  is a fuzzy measure.

**Example 5.** We recall the Example 4 in order to show the calculation of  $\mu_{P^+}$  for a given set  $V$  with 4 elements. We consider the membership functions  $\eta_{X_A}$  and  $\eta_{X_B}$  defined in Table 3. We consider the functions  $\varphi = \max$ ,  $\phi = \text{product}$  and  $N(x) = 1 - x$ .

**Table 3.** Membership degree of each element of  $V$  to the poles  $X_A$  and  $X_B$ .

Element	$\eta_{X_A}$	$\eta_{X_B}$
1	1	0
2	0	1
3	1	0
4	0	1

Note that  $\widetilde{JDJ}(V) = 2$  is the amount of peaceful dialogues between the 4 elements. These dialogues come from the groups  $\{1, 3\}$  and  $\{2, 4\}$ .

**Table 4.** Values of the fuzzy measures  $\mu_{P^-}$  and  $\mu_{P^+}$ .

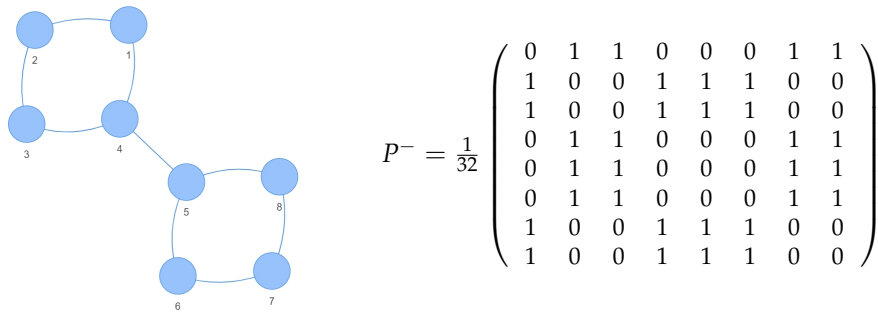
S	$\{1, 2\}$	$\{1, 3\}$	$\{1, 4\}$	$\{2, 3\}$	$\{2, 4\}$	$\{3, 4\}$	$\{1, 2, 3\}$	$\{1, 2, 4\}$	$\{1, 3, 4\}$	$\{2, 3, 4\}$	$\{1, 2, 3, 4\}$
$\mu_{P^-}(S)$	0.25	0	0.25	0.25	0	0.25	0.5	0.5	0.5	0.5	1
$\mu_{P^+}(S)$	0	0.5	0	0	0.5	0	0.5	0.5	0.5	0.5	1

Once we have defined two opposite models to represent the capacity of a set of elements to argue/dialogue, we define a new representation model: the polarization extended fuzzy graph. It combines the ability of a crisp graph to represent a set of elements connected to each other, with the representation of the synergies between these elements, regardless their connections. Hence, from a crisp graph, two membership functions and two aggregation operators, we can define a polarization extended fuzzy graph, a tool which sets light on the modeling of reality.

**Definition 10** (Polarization extended fuzzy graph). Let  $G = (V, E)$  denote a crisp graph, whose nodes set is  $V$  and whose edges set is  $E$ . Let  $\eta_{X_A}$  and  $\eta_{X_B}$  denote the membership functions of the elements of  $V$  concerning the extreme poles  $X_A$  and  $X_B$ . Let functions  $\varphi : [0, 1]^2 \rightarrow [0, 1]$  and  $\phi : [0, 1]^2 \rightarrow [0, 1]$  denote a grouping and a conjunction operator, respectively. Let  $\mu_{P^-} : 2^V \rightarrow [0, 1]$  according denote the fuzzy measure characterized in the equation (3). Then, the triplet  $\tilde{G} = (V, E, \mu_{P^-})$  is a polarization extended fuzzy graph.

Note that the representation ability of the polarization extended fuzzy graph goes far from the modeling provided by other tools as for example, a fuzzy graph. Let us show a toy example.

**Example 6.** We consider the graph  $G = (V, E)$  shown in the Figure. Let the membership functions  $(\eta_{X_A}(1), \dots, \eta_{X_A}(8)) = (1, 0, 0, 1, 1, 1, 0, 0)$  and  $(\eta_{X_B}(1), \dots, \eta_{X_B}(8)) = (0, 1, 1, 0, 0, 0, 1, 1)$  define the membership degree of each element of  $V$  to the poles  $X_A$  and  $X_B$  respectively. We consider  $\varphi = \max$  and  $\phi = \text{product}$ . In Figure 4 we show a representation of the polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_{P^-})$ .



**Figure 4.** Polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_{P^-})$

It may seem that the polarization extended fuzzy graph has a weak point, related to the high complexity concerning the definition of the corresponding fuzzy measure  $\mu_{P^-}$ . Nevertheless, we will show some desirable properties of it, which facilitate the handling of  $\tilde{G}$ . The most important is about the additivity, as it is shown below.

**Proposition 2.**  $\mu_{P^-}$  is a 2-additive fuzzy measure.

**Proof.** We base this demonstration on an asseveration found in [35], where Grabisch demonstrated that a fuzzy measure  $\mu$  is 2-additive if and only if, for all  $S \subseteq V$ , it can be defined as a linear combination  $\mu(S) = \sum_{i=1}^n a_i x_i + \sum_{\{i,j\} \subset S} a_{ij} x_i x_j$ , where  $x_i = 1$  if  $i \in S$ , and  $x_i = 0$  otherwise.

For every  $i \in V$ , we define  $a_i = 0$ , and for every  $i, j \in V$  such that  $i \neq j$ , we define  $a_{ij} = P_{ij}^-$ . Then, according to the Equation (3),  $\mu_{P^-}(S) = \sum_{i,j \in S} P_{ij}^- = \sum_{i,j \in V} P_{ij}^- x_i x_j = \sum_{i=1}^n a_i + \sum_{i,j \in V} a_{ij} x_i x_j$ .  $\square$

**Proposition 3.**  $\mu_{P^-}$  is closed for convex linear combinations:

$$\mu(\sum_{k=1}^m \alpha_k \hat{P}^{-k}) = \sum_{k=1}^m \alpha_k \mu_{\hat{P}^{-k}}$$

where  $\sum_{k=1}^m \hat{P}^{-k} = P^{-k}$  and  $\sum_{k=1}^m \alpha_k = 1$ .

**Proof.** According to Equation (3), and assuming that  $x_i = 1$  if  $i \in S$ , and  $x_i = 0$  otherwise, we have:

$$\begin{aligned} \mu(\sum_{k=1}^m \alpha_k \hat{P}^{-k}) &= \sum_{i,j \in S} \left( \sum_{k=1}^m \alpha_k \hat{P}_{ij}^{-k} \right) x_i x_j = \sum_{k=1}^m \sum_{i,j \in S} \alpha_k \hat{P}_{ij}^{-k} x_i x_j = \\ &= \sum_{k=1}^m \alpha_k \sum_{i,j \in S} \hat{P}_{ij}^{-k} x_i x_j = \sum_{k=1}^m \alpha_k \mu_{\hat{P}^{-k}} \end{aligned}$$

$\square$

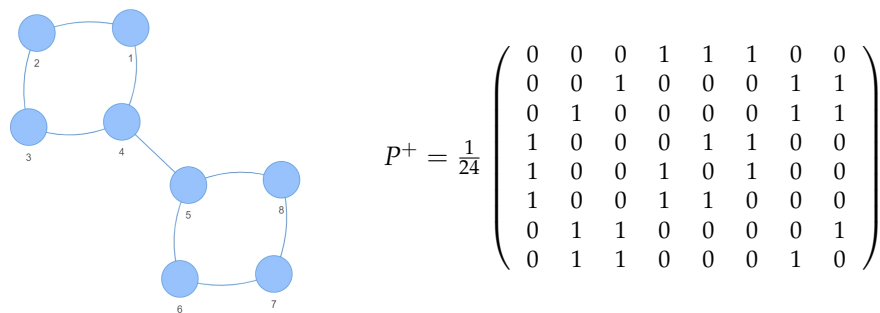
**Remark 6.** As a particular case of the Proposition 3, it holds that  $\mu_P$  is fixed for the mean as follows:

$$\frac{\mu_{P-1} + \mu_{P-2}}{2} = \mu_{\frac{P-1+P-2}{2}}$$

It is worth highlighting that all the points and properties enunciated with respect to  $\mu_{P-}$  also apply to  $\mu_{P+}$ . Then,  $\mu_{P+}$  is a 2-additive fuzzy measure. Particularly, we emphasize in the definition of the non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_{P+})$ , concerning a crisp graph and a non-polarization fuzzy measure.

**Definition 11** (Non-polarization extended fuzzy graph). Let  $G = (V, E)$  denote a crisp graph, whose nodes set is  $V$  and whose edges set is  $E$ . Given a uni-dimensional variable  $X$  with two extreme poles  $X_A$  and  $X_B$ , let  $\mu_{P+}$  the non-polarization fuzzy measure characterized in the Definition 9. Then, the triplet  $\tilde{G} = (V, E, \mu_{P+})$  is a non-polarization extended fuzzy graph.

**Example 7.** We recall the Example 6, but in this case we focus on the measure  $\mu_{P+}$ . Hence, we have the graph  $G = (V, E)$  and membership functions  $(\eta_{X_A}(1), \dots, \eta_{X_A}(8)) = (1, 0, 0, 1, 1, 1, 0, 0)$  and  $(\eta_{X_B}(1), \dots, \eta_{X_B}(8)) = (0, 1, 1, 0, 0, 0, 1, 1)$ . We consider  $\phi = \max$  and  $\phi = \text{product}$ , and  $N(x) = 1 - x$ . Then, the non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_{P+})$  is shown in Figure 5, in which we show structure of the crisp graph and the matrix  $P^+$  concerning  $\mu_{P+}$ .



**Figure 5.** Non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_{P+})$

Note that the measure  $JDJ_{pol}$  quantifies the distance or discrepancy between all pairs of elements  $i, j$  of a given set of individuals  $V$ , i.e., the risk of polarization. Hence, its negation  $\widetilde{JDJ}_{pol}$  can be understood as the minimum risk of polarization for a given population or community. On this assumption, if we consider the non-polarization fuzzy measure  $\mu_{P+}$ , grouping nodes according to this criterion allows us to build communities with minimum risk of conflict. These communities detected in the non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_{P+})$  present a structure that better fixes reality than those communities built only by the relations between nodes.

As the aim of this paper is related to community detection problems, hereafter we will focus on the non-polarization fuzzy measure  $\mu_{P+}$  and the corresponding matrix  $P^+$ . Both tools allow us to manage the synergies between the nodes. To simplify the notation, we consider  $\mu_P = \mu_{P+}$  and  $P = P^+$ .

#### 4. A parametric approach to community detection problem based on Polarization measures and weighted mean

Many complex networks show a modular structure so that the individuals are organized into modules with dense internal connections. Numerous examples can be found: in the field of social networks, groups of related users according to their interests or background; in any citation network, groups of connected papers concerning one particular issue; in a recommendation network, set of

similar services or offers; in metabolic networks, connected biochemical pathways [7,36,37]. Due to the increasingly demand for all these real-life applications among many others, having a consistent community structure helps to understand the main characteristics, functions and topology of these systems. So that, a good understanding of the community structure hidden in a complex network may be helpful for better analysis and exploitation of the data in an effective way [38,39].

Complex networks are usually represented by graphs. One of their most popular applications is devoted to the resolution of community detection problems, whose main goal is to find a *good* partition of a given network. A partition of a graph  $G = (V, E)$  is a decomposition of the set of nodes  $V$  into subgroups known as communities or clusters whose composition depends on the similarity between the objects considered, i.e. a division of the set of nodes into groups that are densely intra-connected, whereas sparsely connected with the rest of the graph [40–42].

Classical algorithms proposed in this field are based on topological information and on the structure of the network considered. Nevertheless, it is undeniable that in the process of modeling the reality by means of a network for subsequent groups search, there is lot of knowledge and information that are not considered in the grouping process. Several authors agree on the importance of adding some additional information to the structure represented by a graph to enrich the communities detected [9,11,43,44]. In this work, it is particularly interesting the problem addressed by Gutiérrez et al. [13–15,45–48] about the detection of communities in extended fuzzy graphs. They proposed a methodology to analyze independently the structural information of the graph and the knowledge represented by the fuzzy measure when grouping the nodes.

We approach the community detection problem based on fuzzy measures including this information about the relations among the individuals emanating from that knowledge about the respective positions in any attitudinal axis. These relations will be considered in terms of Polarization measures built from the  $JDJ_{pol}$  measure. Then, the base of the problem here addressed is a non-polarization extended fuzzy graph  $\tilde{G} = (G = (V, E), \mu_P)$ . It is worth highlighting the increment of cohesiveness procured in the groups by considering additional information independent of the topology. Note that we consider the non-polarization fuzzy measure  $\mu_P = \mu_{P+}$  instead of the polarization fuzzy measure  $\mu_{P-}$  in order to fix an scenario in which all the components of  $\tilde{G}$ ,  $A$  and  $\mu_P$ , have the same somehow ‘positive’ nature.

To face this problem, we work inspired by the idea developed in the Additional Louvain algorithm (see [14]), based on the Louvain algorithm [8]. The key point is to distinguish two different roles within the input parameters: one of them, to establish the neighbour relations, and the other, to calculate the variation of the modularity. The first role will be played by the adjacency matrix of the graph,  $A$ , so that only those nodes that are connected in  $G$  can be in the same group. On the other hand, we suggest to consider a combination of the two components of the non-polarization extended fuzzy graph  $\tilde{G}$  as basis to calculate the variation of modularity, in order to incorporate the additional information. Then, having a crisp graph  $G$ , the two membership functions  $\eta_{X_A}$  and  $\eta_{X_B}$ , the operators  $\varphi$  (grouping) and  $\phi$  (overlapping) and considering the negation function  $N$ , or what is the same, a non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_P)$ , we propose a new methodology that is summarized as follows.

1. Obtain the non-polarization fuzzy measure  $\mu_P$  related to the set  $V$  from the parameters  $\eta_{X_A}$ ,  $\eta_{X_B}$ ,  $\varphi$ ,  $\phi$  and  $N$ , according to equation (8).
2. Summarize  $\mu_P$  into a matrix,  $F$ .
3. Define the matrix  $M = \theta(A, F)$ , where  $\theta : \Pi(n)^2 \rightarrow \Pi(n)$  is a matrix aggregator used to combine two matrices into a single one.
4. Apply the Louvain algorithm by distinguishing the role of the matrices  $A$  and  $M$ :  $A$  is used to find the neighbour relations,  $M$  is used to calculate the variation of modularity.

**Remark 7.** In this proposal we suggest the use of a matrix aggregator  $\theta$ . Nevertheless, any other operator could be applied instead.

The definition of the matrix  $F$  as an aggregation of the non-polarization fuzzy measure  $\mu_P$ , should be closely related to the problem addressed. We suggest a particular characterization of it, based on the calculation of the weighted graph associated with  $\mu_P$ . This matrix is a highly recommended tool for fuzzy measures manipulation and visualization, which summarizes the knowledge about the capacity of the elements into  $n^2$  data set. The definition of this graph is based on the Shapley value [49], particularly in its characterization related to fuzzy measures [50].

**Definition 12** (Weighted graph associated with a fuzzy measure  $G_\mu$  [13,14]). Let  $\mu : 2^V \rightarrow [0, 1]$  denote a fuzzy measure defined over the finite set  $V$ , and let  $\xi : [-1, 1]^2 \rightarrow [0, 1]$  denote a bi-variate aggregation operator. We consider  $Sh_i(\mu)$ , the Shapley value of the individual  $i \in V$  in coalition with all the elements of  $V$  regarding their relation in  $\mu$ ; analogously,  $Sh_i^j(\mu)$  denotes the Shapley value of the individual  $i$  in coalition with all the elements of  $V \setminus \{j\}$ , regarding  $\mu$ . Then, the weighted graph associated with the fuzzy measure  $\mu$ , denoted by  $G_\mu$  is that whose adjacency is represented by the matrix  $F$ , where:

$$F_{ij} = \xi \left( Sh_i(\mu) - Sh_i^j(\mu), Sh_j(\mu) - Sh_j^i(\mu) \right) \quad (9)$$

In our specific proposal of the method to find communities in a non-polarization extended fuzzy graph, we suggest summarizing the non-polarization fuzzy measure  $\mu_P$  into the matrix  $F$ , adjacency of its associated weighted graph. To formally establish this method, let us define it as an algorithm, named Polarization Louvain, whose pseudocode can be found in Algorithm 1.

The key point to approach a clustering process in a non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_P)$  is the calculation of the weighted graph associated with  $\mu_P$ . The calculation of the Shapley value on which it is based, is a process which usually reaches exponential complexity. Nevertheless, we will show that this problem does not apply when considering  $\mu_P$ , for which we have demonstrated in Proposition 2 that it is a 2-additive fuzzy measure.

**Proposition 4.** Let  $\mu_P$  denote the non-polarization fuzzy measure related to set  $V$  obtained from the membership functions  $\eta_{X_A}$  and  $\eta_{X_B}$ , the aggregation operators  $\varphi$  and  $\phi$  and the negation operator  $N$ , according to Definition 9, so that  $P$  is the matrix obtained from these parameters according to equation (7). The following holds for  $\mu_P$ .

1.

$$Sh_i(\mu_P) = \sum_{k \in V} P_{ik}$$

2.

$$Sh_i^j(\mu_P) = \sum_{k \in V \setminus \{j\}} P_{ik} = \left( \sum_{k \in V} P_{ik} \right) - P_{ij}$$

**Proof.** We prove the point 1, so that the demonstration of 2 is analogous.

It is based on an alternative characterization of the Shapley value [51,52] in which,  $\forall i \in V$ , the corresponding Shapley index can be calculated as the average of the marginal contributions in all the permutations of the original set  $V$ , i.e.,

$$Sh_i = \frac{1}{n!} \sum_{o \in \pi(n)} [\mu(\text{pred}(i) + \{i\}) - \mu(\text{pred}(i))]$$

where  $\text{pred}(i)$  denotes the set of predecessors of  $i$  in the order  $o$  and  $\pi(n)$  denotes the set of all the possible permutations of a set with  $n$  elements.

According to Equation (8),

$$\mu(\text{pred}(i) + \{i\}) = \sum_{k=1}^n \sum_{j=1}^n P_{jk} x_j x_k + \sum_{k=1}^n (P_{ik} + P_{ki}) x_k$$

$$\mu(pred(i)) = \sum_{k=1}^n \sum_{j=1}^n P_{jk} x_j x_k$$

being  $x_j = 1$  if  $j \in pred(i)$ , and  $x_j = 0$  otherwise.

So that,

$$Sh_i = \frac{1}{n!} \sum_{o \in \pi(n)} \left( \sum_{k=1}^n \sum_{j=1}^n P_{jk} x_j x_k + \sum_{k=1}^n (P_{ik} + P_{ki}) x_k \right) - \left( \sum_{k=1}^n \sum_{j=1}^n P_{jk} x_j x_k \right) = \frac{1}{n!} \sum_{o \in \pi(n)} \sum_{k=1}^n (P_{ik} + P_{ki}) x_k$$

For a half of the orders  $o \in \pi(n)$ , it is true that  $k \in pred(i)$ , so, for a half of the values of the previous summation,  $x_k = 1$ . Therefore,

$$\frac{1}{n!} \sum_{o \in \pi(n)} \sum_{k=1}^n (P_{ik} + P_{ki}) x_k = \frac{1}{2} \sum_{k=1}^n (P_{ik} + P_{ki})$$

By definition,  $P$  is symmetric, and its main diagonal is null. Hence,  $P_{ik} = P_{ki}$  and  $P_{ii} = 0$ . Then,

$$\frac{1}{2} \sum_{k=1}^n (P_{ik} + P_{ki}) = \frac{1}{2} \sum_{k=1}^n 2P_{ik} = \sum_{k=1}^n P_{ik} = \sum_{k \in V} P_{ik}$$

□

As a consequence of the Proposition 4, the following result holds for  $\mu_P$ .

**Proposition 5.** Let  $\mu_P$  denote the non-polarization fuzzy measure related to set  $V$  obtained from the membership functions  $\eta_{X_A}$  and  $\eta_{X_B}$ , the aggregation operators  $\varphi$  and  $\phi$  and the negation operator  $N$ , according to Definition 9, so that  $P$  is the matrix obtained from these parameters according to equation (7). Let  $i, j \in V$  denote two individuals. Then, the following applies.

$$1. Sh_i(\mu_P) - Sh_i^j(\mu_P) = P_{ij}$$

$$2. Sh_j(\mu_P) - Sh_j^i(\mu_P) = P_{ji}$$

**Proof.** We prove the point 1, so that the demonstration of 2 is analogous.

As it is demonstrated in the Proposition 4,

$$1. Sh_i(\mu_P) - Sh_i^j(\mu_P) = \sum_{k \in V} P_{ik} - \left( \sum_{k \in V} P_{ik} - P_{ij} \right) = \sum_{k \in V} P_{ik} - \sum_{k \in V} P_{ik} + P_{ij} = P_{ij}$$

□

At this point, it is trivial to represent the closeness between different pairs of elements according to their attitude concerning a particular issue. Then, being  $\mu_P$  the corresponding non-polarization fuzzy measure based on the Polarization measure  $JDJ_{pol}$ , we assume that the closeness between two individuals concerning its attitude about one issue can be represented by the weighted graph associated with  $\mu_P$ , i.e., with the corresponding adjacency matrix of  $G_{\mu_P}$ , calculated as:

$$F_{ij} = \xi \left( Sh_i(\mu_P) - Sh_i^j(\mu_P), Sh_j(\mu_P) - Sh_j^i(\mu_P) \right) = \xi(P_{ij}, P_{ji})$$

**Remark 8.** Note that because of  $P$ 's symmetry, if the chosen aggregation operator  $\phi$  is of the type  $\max$ ,  $\min$ , average among others, then  $F_{ij} = P_{ij}$ ,  $\forall i, j \in V$ .

So far, we have summarized all the knowledge modeled by the non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_P)$  into two independent matrices,  $A$  and  $F$ . This process/tools could be applied in many fields, as for example problems about centrality, link prediction or propagation.

It is crucial to be clear about the interpretation of the matrices  $A$  and  $F$  (or  $P$  in any case). On the one hand,  $A$  represents the direct connections between the elements of  $V$ ; it is well accepted that nodes tightly-knit connected should be connected, so it can be seen that the connections shown in  $A$  are 'positive'. On the other hand, we have already mentioned that, because of the characterization of  $\mu_P$  (and hence of  $P/F$ ), it is related to the synergies or closeness between the elements. Then, we can conclude that both matrices  $A$  and  $F$  have 'positive' meanings, so that nodes for which both matrices (or even one of them if it is fair enough) define high values, should be together.

Let us illustrate the performance of the Polarization Louvain method with a toy example. In this case, we combine the matrices  $A$  and  $F$  by means of a linear combination  $\theta(A, F) = \gamma A + (1 - \gamma)F$ . In our opinion, it is an smart way to assign a weight or importance to each component of the  $\tilde{G}$ . Note that, when  $\gamma = 1$ , the additional information is not considered. In this case, both the search of neighbour relations and the modularity variation are calculated over the matrix  $A$ , so that the Polarization Louvain algorithm is exactly the same than the Louvain algorithm.

**Example 8.** Let us consider the graph  $G = (V, E)$  whose adjacency matrix is  $A$ ; and let us assume some knowledge about the position of the elements of  $V$  in any attitudinal axis modeled, in the sense that we know the membership degree of all the individual in  $V$  to the poles  $X_A$  and  $X_B$ , represented by the membership functions  $\eta_{X_A}$  and  $\eta_{X_B}$  respectively. From this knowledge, and considering the operators  $\varphi = \max$  and  $\phi = \text{prod}$  and  $N(x) = 1 - x$ , we define the fuzzy measure  $\mu_P$ , and hence the matrix  $P$  (equation (8)).

**Table 5.** Membership degree of each node of  $V$  to the poles  $X_A$  and  $X_B$ .

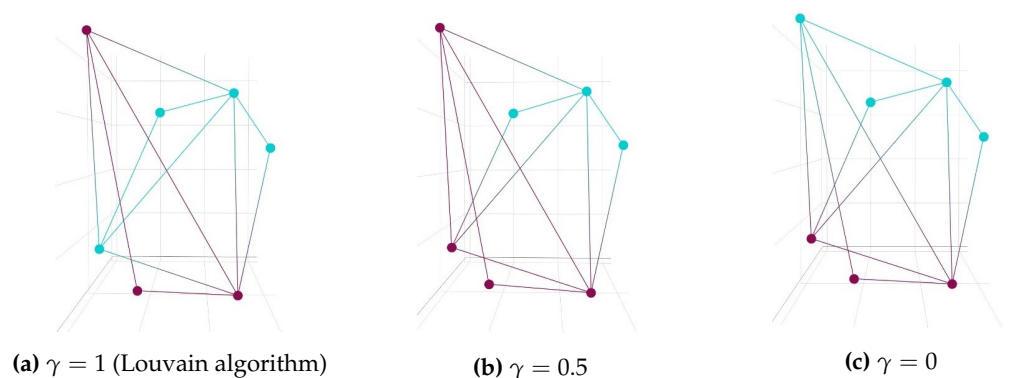
Vertex	$\eta_{X_A}$	$\eta_{X_B}$
1	0.022	0.878
2	0.756	0.144
3	0.751	0.099
4	0.5	0.5
5	0.001	0.989
6	0.102	0.888
7	0.889	0.112

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \end{pmatrix} \quad P = \frac{1}{22.574} \begin{pmatrix} 0.000 & 0.336 & 0.341 & 0.561 & 0.978 & 0.910 & 0.219 \\ 0.336 & 0.000 & 0.892 & 0.622 & 0.252 & 0.329 & 0.872 \\ 0.341 & 0.892 & 0.000 & 0.625 & 0.257 & 0.333 & 0.912 \\ 0.561 & 0.622 & 0.625 & 0.000 & 0.506 & 0.556 & 0.556 \\ 0.978 & 0.252 & 0.257 & 0.506 & 0.000 & 0.899 & 0.121 \\ 0.910 & 0.329 & 0.333 & 0.556 & 0.899 & 0.000 & 0.211 \\ 0.219 & 0.872 & 0.912 & 0.556 & 0.121 & 0.211 & 0.000 \end{pmatrix}$$

**Figure 6.** Matrices  $A$  and  $P$

Due to the properties of  $P$ , we have that  $F = P$ , so the characterization of  $\mu_P$  is straightforward.

At this point, we have the non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_P)$ , so that, being  $\gamma \in [0, 1]$  a balancing factor, and considering the aggregation function  $\theta(A, F) = \gamma A + (1 - \gamma)F = \gamma A + (1 - \gamma)P$ , we apply the Polarization Louvain algorithm. In the Figure 7 we show the partitions obtained for several values of  $\gamma$ . Note how the way in which the nodes are organized changes depending on the importance assigned to the information represented by  $P$  (i.e.  $F$ ) about the closeness between the nodes.



**Figure 7.** Partitions of  $\tilde{G} = (V, E, \mu)$  obtained with the Polarization Louvain algorithm.

## 5. A real case: the impact of the COVID-19 pandemic in the organization of the people

### 5.1. Experiment design: sources and methodology

We briefly explain the case of study in the following. The nodes and their relations considered in this work have been obtained from the social network Twitter, particularly from some posts recorded along the state of alarm imposed by the central government in Spain (from March 16th, 2020 till June 29th, 2020). All data downloaded is related to the COVID-19 pandemic and the political situation in that country, concerning the management of the sanitary situation by the Spanish government. Each element of that data set represents an influential and verified account.

It is well known by popular knowledge that Twitter is one of the trendiest online social networks, where millions of users debate about any social or political topics, among many others. For this research, we have used the retweet (RT) network. A RT is a post derived from the action of any user to replicate a given tweet or message of other user to spread that content with his/her followers. In the literature, the RT network has been commonly used as a directed network [53], so that if the original tweet is written by an user  $i$  and then  $j$  RT it, then there is a connection with directionality, which represents the action of each user. Nevertheless, in our case we understand it as a non-directed network in this sense: it is not of our matter to know the directionality of a connection (who posts the tweet and who RT it), but rather to focus on the content. In this vein, once a given user  $j$  RT a tweet of  $i$ , what is important to us is the intention of  $j$  to transmit and spread that content. In broader terms we can assume that  $j$  agrees with the content and spreads the word to make the tweet visible and influential over the people. Our aim is to know the user's political attitude towards the Spanish government measuring their attitude reflected on the tweets. Hence, no directionality is needed.

All data was downloaded from Twitter, using its API by R-Studio, with the package "rtweet" [54] in 5 rounds along the state of alarm in Spain.

- ▷ 1st round: "2020-03-16" - "2020-03-23".
- ▷ 2nd round: "2020-04-06" - "2020-04-21".
- ▷ 3rd round: "2020-05-07" - "2020-05-22".
- ▷ 4th round: "2020-06-03" - "2020-06-15".
- ▷ 5th round: "2020-06-14" - "2020-06-29".

The criteria we used to download the tweets was related to the considerations of those keywords which are mainly composed by the main political parties in Spain as well as their leaders:

*psoe OR pp OR vox OR ciudadanos OR gobierno OR podemos OR españa OR sanchezcastejon OR vox\_es OR pabloiglesias OR pablocasado OR santi\_abascal OR inesarrimadas OR CiudadanosCs OR populares OR estadodealarma*

After the downloading phase, we obtained 4895747 tweets. Then, it was needed a manual encoding applied in a sample of those tweets to fix the points:

- (1) To detect and filter all those tweets included on our database which do not correspond to our goals (Feature: **TOPIC**).
- (2) To encode each tweet as (a) detractor, (b) neutral or (c) supporter of the Spanish government (Feature: **POSITION**).

The manual encoding was applied to a random sample of 1500 tweets for each round mentioned above. It is worth mentioning the importance of encoding by rounds due to the dynamic nature of debates on online social networks, in which words or events can change over time despite being debating about one specific topic. Once the data was encoded, we applied text classification with machine learning algorithms in order to tackle the full content of our database.

According to [55], *linear Support Vector Machines* are recognized to be one of the best machine learning algorithms for text classification. So that, after the *tokenization* phase and the removal of *stopwords*, we convert our text into a *tf-idf* matrix. This type of matrices present all the different words which appeared on the corpus on the columns, and the strings (tweets in our case) on the rows. The simplest *dfm* matrix is an occurrence matrix with 0 if a given word does not appear on the tweet and 1 if it does. However, *tf-idf* matrices show values as a result of the product of a term frequency and inverse document frequency for each word of a tweet. So that, as it is a classification problem with the classes *detractor* (pole  $X_A$ ) and *supporter* (pole  $X_B$ ), the classifier was trained and applied for the feature "TOPIC" and then for the "POSITION". The results obtained with that process of text classification are showed in the Table 6. Note that the final scores recorded for "POSITION" are the two probabilities for being a "detractor" or "supporter" tweet towards the Spanish government. In this case, "neutral" category is omitted in order to get a variable with two poles, assuming that probabilities close to 0.5 correspond to the "neutral" category.

In the Table 6 we show an analysis in which the following indexes have been considered: *PRECISIÓN*, *RECALL*, *KAPPA*, *F-SCORE* y *AUC* [56,57].

**Table 6.** Linear SVM performance for features "TOPIC" and "POSITION".

Round	Feature	Precision	Recall	Kappa	F-Score	AUC
1	TOPIC	0.8017	0.9322	0.3670	0.8620	0.6583
2	TOPIC	0.8167	0.5476	0.5077	0.6556	0.7344
3	TOPIC	0.8267	0.7027	0.6187	0.7596	0.8010
4	TOPIC	0.7867	0.7090	0.564	0.7457	0.7791
5	TOPIC	0.7659	0.8758	0.5216	0.8171	0.7567
1	POSITION	0.8492	0.9854	0.4816	0.9122	0.6950
2	POSITION	0.8960	0.9619	0.7761	0.9277	0.8780
3	POSITION	0.8392	0.8488	0.6675	0.8439	0.8366
4	POSITION	0.9133	0.9048	0.8225	0.9090	0.9121
5	POSITION	0.8318	0.8600	0.6638	0.8456	0.8335

Finally, the database derived from the SVM classifier is integrated by 1208631 tweets and 469616 users. To aim for those *influencers* and verified accounts, we filtered by

- (a) Tweets with high repercussion on Twitter, considering accounts whose tweets with RT count are placed above the 50 percentile ( $n \geq 317$ ).
- (b) Verified accounts.
- (c) Accounts with high number of followers, considering accounts whose number of followers is placed above the 50 percentile ( $n \geq 21779$ ).

In this manner, 406 users left which are mainly politicians, party accounts and journalists. Then, to get a closed network of users, we matched those accounts that both write or RT any tweet of those 406

accounts. So that, in our final data base, 295 users are considered, from whose posts, 657 interactions are derived. Note that these interactions may concern users who are not among the 295 considered, but who have RT some of theirs posts; so that we have a total amount of 454 different users and 657 interactions. From this information, we build a network  $G = (V, E)$ , so that the each of these 454 accounts is represented by a node of the set  $V$ , and the links represents the edges of  $E$ . Let us remark that we take into account if it comes the case in which two users interact several times (by means of RT of different tweets), i.e. we work with a weighted graph, so that weight  $w_{ij}$  of the corresponding edge represents how many time have interacted the users  $i$  and  $j$ .

Note that the objects to be classified were not users but tweets, so, for each user, we computed the average score for his/her tweets of being "detractors" and "supporters" (not only considering the original posts, but also the RT). At this end, we finally got, for the 295 users, two specific values of probability for being a "detractor" and for being a "supporters" toward the Spanish government. These distances to the support vector machines for each class of the SMV, will be used as membership degree values for each user to compute the  $JDJ_{pol}$  Polarization measure, which sets the basis of the non-polarization measure  $\mu_P$  (Definition 9) which is part of the non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_P)$  over which we apply the community detection problem.

For a better understanding of the results, it is important to provide a proper visualization of the network, which comprises a complex process [58]. Having a proper organization and a good representation of the network itself is fundamental for a better understanding and exploitation of the inherent data. To accomplish that, we have used the R package "visNetwork" [59].

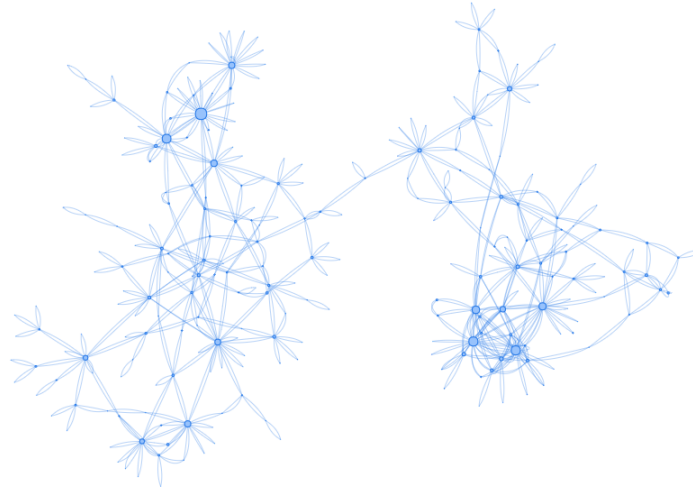
## 5.2. Results

In this section we show the computational results obtained when applying the Polarization Louvain algorithm to the data set obtained from Twitter as explained in Section 5.1. To carry on with it, we build a graph from that obtained data set and a fuzzy measure which represents the capacity of the elements to trigger conflict. Originally, this graph had 454 nodes and 657 weighted edges (the list of the interactions between users from which we define the set of edges can be found in GitHub<sup>1</sup>. Nevertheless, for the clustering process, we focus on its weak component, which contains 261 nodes and 484 weighted edges. The obtained network,  $G = (V, E)$  with adjacency matrix  $A$ , is showed in the Figure 8 (taking into consideration the weight of each edge, the degree of the nodes is represented by their size in the image, so that the bigger nodes will represent the users with the most amount of interactions). Then, considering the membership degrees of each node to the poles  $X_A$  (being a "detractor" of the Spanish government) and  $X_b$  (being a "supporter" of the Spanish government), represented by  $\eta_{X_A}$  and  $\eta_{X_B}$  respectively, we can calculate the Polarization measure  $JDJ_{pol}$  (see Definition 7) from which we define the matrix  $P$ , according to the equation (7). The membership degrees considered can be found in GitHub<sup>1</sup>. It provides us the non-polarization fuzzy measure  $\mu_P$  which is one of the components of  $\tilde{G} = (V, E, \mu_P)$ .

The measure  $\mu_P$  depends on the selection of a negation operator,  $N$ , and two different types of aggregation operators: a grouping function  $\varphi$  and an overlapping operator  $\phi$ . As negation operator, we use  $N(x) = (1 - x)$ . Concerning the aggregation operators, we use some of the most important operator in this field, having two different scenarios for the aggregation of the membership degrees: (a)  $\phi = \min$  and  $\varphi = \max$ ; (b)  $\phi = \text{product}$  and  $\varphi = \max$ .

Because of the characterization of  $P$ , and being  $\mu_P$  a fuzzy measure characterized as in equation (8),  $P$  can be seen also as the adjacency matrix of  $G_{\mu_P}$ , so we can indistinctly consider both tools.

<sup>1</sup> <https://github.com/inmaggp/Community-Detection-Problem-Based-on-Polarization-Measures.-An-application-to-Twitter-the-COVID-19->



**Figure 8.** Graph  $G = (V, E)$ .

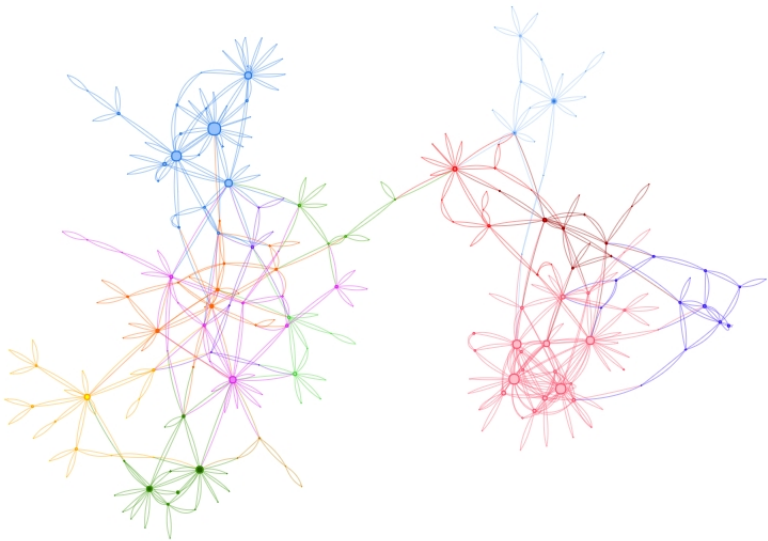
We apply the Polarization Louvain algorithm to find communities in the non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_P)$ . Let us note that the obtained communities will be cohesive with the whole knowledge modeled by it, the structure of the graph as well as the additional information modeled by  $\mu_P$ . The notion of *what is a community* will be closely connected with the aggregation operator  $\theta$  chosen, as well as with the grouping operator  $\varphi$  and the overlapping operator  $\phi$ .

To combine the two components of  $\tilde{G}$ , we work with linear combinations of the matrices  $A$  and  $P$  assigning them an importance by means of the balancing parameter  $\gamma \in [0, 1]$ , i.e., we consider the matrix  $M = \theta(A, P) = \gamma A + (1 - \gamma)P$ .

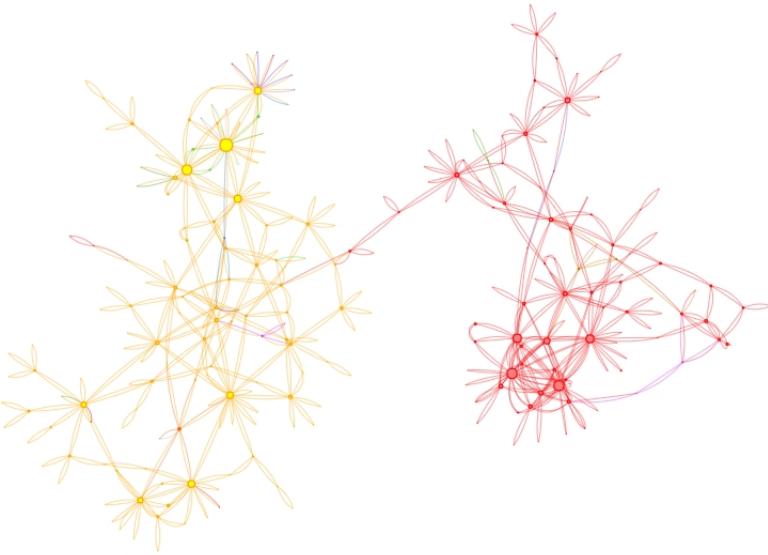
The influence of each component of  $\tilde{G}$  varies depending on the value of  $\gamma$ . For values of  $\gamma$  close to 1, the structural component gains importance, so the groups contain nodes tightly-knit connected in  $A$ . On the opposite, when  $\gamma$  is close to 0, the additional information modeled by  $\mu_P$  turns decisive in the definition of the communities, so, if it is possible regarding the structure of  $A$ , the groups contain nodes with low Polarization level, i.e. nodes whose membership degree to each pole is similar. In this case, those users about whom we can assume similar political viewpoint, will be in the same group.

We apply the Polarization Louvain algorithm for the two scenarios of grouping/overlapping functions previously mentioned, and considering the matrix  $M = \gamma A + (1 - \gamma)P$ , for several values of the importance parameter,  $\gamma = 0.5, 0.4, 0.3, 0.2, 0.1, 0$ . We also compute the Louvain algorithm with matrix  $A$ , on whose result is based our comparison analysis. Note that the performance of the Louvain algorithm matches with the Polarization Louvain algorithm when  $M = A$ , ( $\gamma = 1$ ).

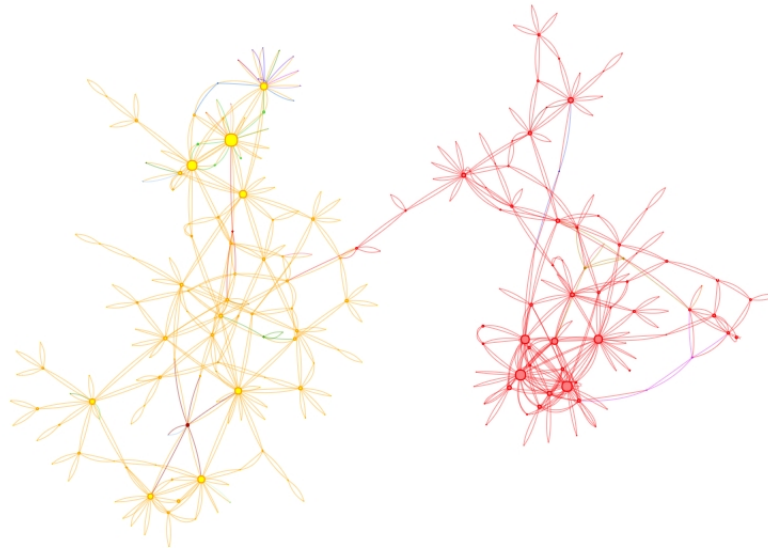
Here we show how the organization of the groups keep changing depending on the importance of each component of  $\tilde{G}$  in the clustering process. Particularly, for the extreme cases, Louvain (in which there is no additional information), and  $\gamma = 0$  (in which the additional information gains all the importance), considering the two scenarios previously mentioned about the aggregation operators used. In GitHub, we include a file in which we show the obtained partitions for every value of  $\gamma$  considered, as well as the corresponding images.



**Figure 9.** Partition obtained with the Louvain algorithm in the graph  $G = (V, E)$ .



**Figure 10.** Partitions obtained with the Polarization Louvain algorithm in the non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_p)$ .  $\gamma = 0$ ;  $\varphi = \max$ ;  $\phi = \min$ .



**Figure 11.** Partitions obtained with the Polarization Louvain algorithm in the non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_P)$ .  $\gamma = 0$ ;  $\varphi = \max$ ;  $\phi = \text{prod}$ .

Note how, when only the political viewpoint of the users is considered, the graph is divided into two main communities, so that we can easily differentiate between the detractors and the supporters of the Spanish government.

To measure the goodness of the obtained partitions, we refer to the  $JDJ_{pol}$  measure. We agree on that a cohesive group should be composed by connected users with similar viewpoints. In this sense, we can say that a group is as cohesive as low is its corresponding  $JDJ_{pol}$  value.

Note that the partitions obtained when consider several values of  $\gamma$  vary in the number of communities. Then, to compare them, we consider the weighted average of the  $JDJ_{pol}(C_i)$  value of all its communities. Thus, we calculate the Polarization value of the partition  $P = \{C_1, \dots, C_s\}$  as follows:

$$pol(P) = \left[ \frac{\sum_{i=1}^s JDJ_{pol}(C_i) * |C_i|}{\sum_{i=1}^s |C_i|} \right]_{|C_i| > 1} \quad (10)$$

It is important to put attention on the fact that only the non isolated communities will be considered to calculate  $pol(P)$ , i.e. groups with more than one element  $|C_i| > 1$ ; it does not make any sense consider how polarized is one element with respect itself.

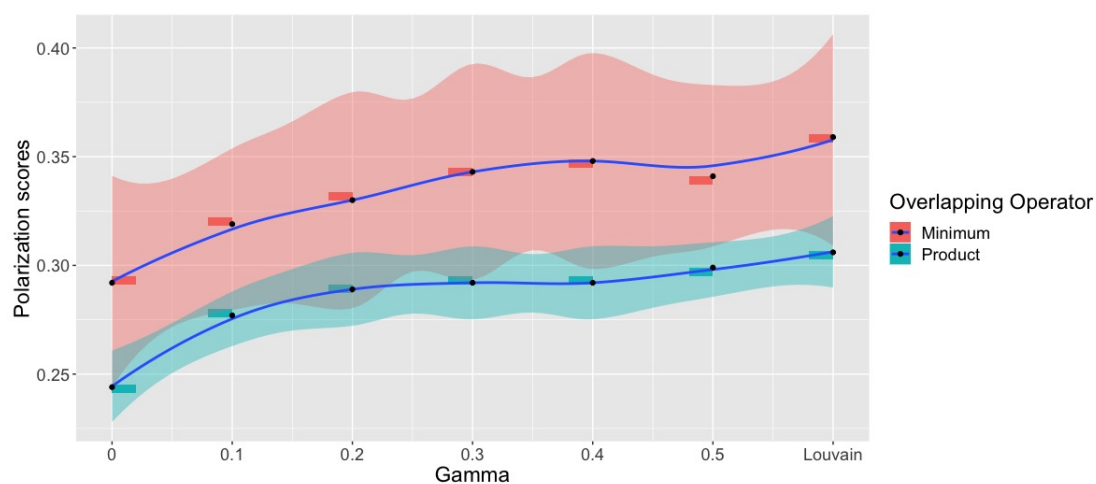
In Tables 7 and 8, we show the  $JDJ_{pol}$  value of each community in the obtained partitions (only non-isolated communities), as well as the corresponding  $pol(P)$ . For each partition, we show the vector  $(JDJ_{pol}(C_1), \dots, JDJ_{pol}(C_s))$ , so that the  $i$ th component corresponds with  $JDJ_{pol}(C_i)$ .

**Table 7.** Comparison of the obtained partitions.  $\varphi = \max$  and  $\phi = \min$ .

$\varphi = \max$ $\phi = \min$	# Communities $ C_i  > 1$	$(JDJ(C_1), \dots, JDJ(C_s))$	$pol(P)$
Louvain	14	(0.256, 0.514, 0.253, 0.301, 0.458, 0.377, 0.302, 0.4403, 0.459, 0.349, 0.190, 0.475, 0.108, 0.415)	0.359
$\gamma = 0.5$	11	(0.239, 0.259, 0.513, 0.297, 0.377, 0.440, 0.514, 0.257, 0.459, 0.415, 0.455)	0.341
$\gamma = 0.4$	8	(0.254, 0.332, 0.259, 0.513, 0.450, 0.514, 0.459, 0.455)	0.348
$\gamma = 0.3$	7	(0.304, 0.300, 0.253, 0.513, 0.512, 0.526, 0.246)	0.343
$\gamma = 0.2$	8	(0.334, 0.267, 0.444, 0.512, 0.462, 0.440, 0.528, 0.246)	0.330
$\gamma = 0.1$	7	(0.323, 0.273, 0.418, 0.482, 0.440, 0.462, 0.246)	0.319
$\gamma = 0$	8	(0.302, 0.263, 0.439, 0.463, 0.277, 0.440, 0.462, 0.246)	0.292

**Table 8.** Comparison of the obtained partitions.  $\varphi = \max$  and  $\phi = prod$ .

$\varphi = \max$ $\phi = prod$	# Communities $ C_i  > 1$	$(JDJ(C_1), \dots, JDJ(C_s))$	$pol(P)$
Louvain	14	(0.218, 0.454, 0.228, 0.261, 0.378, 0.296, 0.261p, 0.359, 0.389, 0.306, 0.168, 0.392, 0.102, 0.258)	0.306
$\gamma = 0.5$	11	(0.220, 0.453, 0.190, 0.260, 0.261, 0.296, 0.359, 0.326, 0.382, 0.389, 0.258)	0.299
$\gamma = 0.4$	9	(0.214, 0.281, 0.220, 0.453, 0.363, 0.389, 0.369, 0.258, 0.343)	0.292
$\gamma = 0.3$	7	(0.257, 0.251, 0.220, 0.453, 0.369, 0.417, 0.343)	0.292
$\gamma = 0.2$	7	(0.259, 0.228, 0.453, 0.369, 0.417, 0.249, 0.186)	0.289
$\gamma = 0.1$	7	(0.274, 0.228, 0.393, 0.417, 0.369, 0.249, 0.186)	0.277
$\gamma = 0$	8	(0.256, 0.224, 0.316, 0.376, 0.199, 0.243, 0.249, 0.186)	0.244

**Figure 12.** Polarization values of the partition  $P = \{C_1, \dots, C_s\}$  by overlapping operators.

As it can be seen in previous tables, as well as in the Figure 12, the  $pol(P)$  value related to those partitions obtained with the Polarization Louvain algorithm is lower than the one related to the partition provided by the Louvain algorithm. Then, we can assert that this method provides more cohesive community structures according to the reality modeled.

## 6. Discussion

There are several points to be discussed at this end.

From a theoretical point of view, it is undeniable that complex models fix better the reality than classical tools. Having several criteria to be considered makes more complex the resolution process of a problem, but it is certainly worth it.

Classically, the methods proposed to find communities in a graph, only analyze its structural features. Far from this assumption, in this work we have taken into consideration several aspects inherent to reality, which, with a proper process of modelling and analysis, can be considered as different criteria in the community detection problem.

Then, we distinguish between different types of information. On the one hand, we deem the crisp knowledge which could be easily related to the classical graphs considered in the literature. This type

of knowledge is unalterable and objective, in the sense that it exists and no changes could be made about it. It is the case of the direct connections represented by the edges of a graph. Particularly, in the real case here addressed based on the Online Social Network Twitter, we have worked with the retweet (RT) network. It is an objective information directly obtained from the social site. On the other hand, we analyze other types of information sources inherent to the people who discuss in Twitter. A wide range of different aspects could be considered, from the factual issues related to objective knowledge about the people as for example the distance that separates them or the common followers, to the more subjective points related to ideology or feelings.

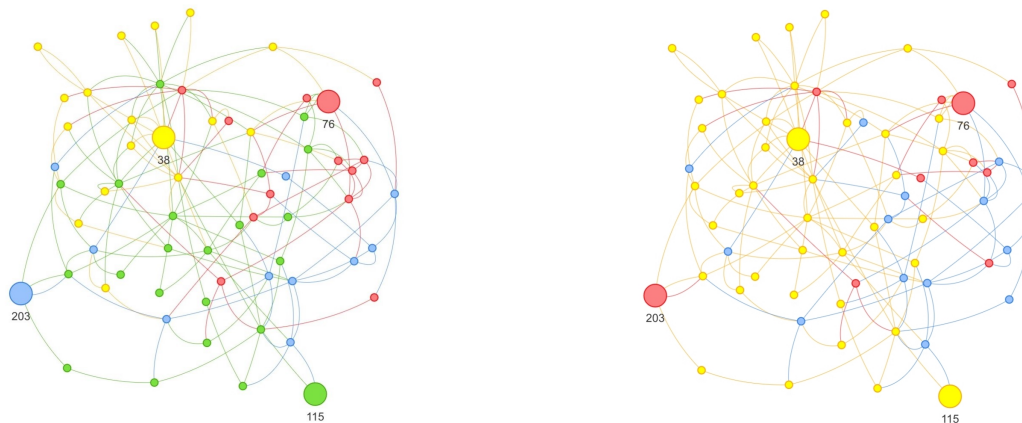
In the context of great political instability enhanced by the global COVID-19 crisis, we agree on analyze the political position of several people who are highly influential on Twitter. The study of feelings, ideology and political principles, is always a hard matter in which many inaccurate details have to be taken into account. To deal with the vagueness and vagueness related to the analysis of the political attitude, we work with fuzzy measures, added in the modeling process to the crisp graph. In this vein, we work with the non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_P)$ , where  $G = (V, E)$  is the weighted graph which represents the RT network, and  $\mu_P$  is a fuzzy measure which defines relations between the elements of  $V$ , depending on their position in a political axis. To define this fuzzy measure, we consider the  $JDJ_{pol}$  measure, which quantifies the Polarization of a given society.

Several criteria have to be fixed for the calculation of  $JDJ_{pol}$ , as for example the aggregation operator  $\phi$  and the grouping function  $\varphi$ . In this paper we have selected considered some of the most popular functions in this field, specifically,  $\varphi = \max$  and  $\phi \in \{\min, product\}$ . Note that these operators play an essential role in the value of the  $JDJ_{pol}$  measure (and also the negation operator  $N$  if we are interested in considering the opposite of  $JDJ_{pol}$ ) and thus in the community detection problem here addressed. So it would be interesting to analyze how the structure of the partitions keeps changing according to the operators considered.

In the same manner, the operator  $\theta$  involved in the Polarization Louvain algorithm impacts on the community structure detected, in terms of how to aggregate both components of a non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_P)$  (the structure and the closeness between the nodes). We agree on considering linear combinations of the two matrices involved, in order to assign an 'importance' to each of them, by means of a balancing or weighting factor  $\gamma \in [0, 1]$ . This procedure allows us to examine the changes that occurs in the structure of communities, according to the how much influence on its definition each of these components. Then, considering the aggregation  $\gamma A + (1 - \gamma)P$ , those values of  $\gamma$  which are close to 1 are related to partitions in which the nodes of the same group are densely connected in  $G$ , whereas for lower values of  $\gamma$  it is important to maintain together nodes with high values of closeness (without omitting the structure of  $G$ ).

Regarding Polarization values, the Louvain algorithm shows the highest values of Polarization, as it can be seen in the Tables 7 and 8, as well as in the Figure 12. In this work, we propose a new method for community detection which in our opinion, has strong theoretical and applied connotations. The extra-information provided by the measure of Polarization  $JDJ_{pol}$  matches up with community detection algorithms due to their close conceptual relationship. The fact of adding Polarization scores implies taking into account the similarity between individuals along an attitudinal axis. In this vein, having new information closely related to the purposes for which the community detection algorithms are applied, makes the communities more cohesive with a greater homogeneity degree, so that this construction of the communities fixes better the reality. In our case, the aim is to cluster the nodes according to their position towards the Spanish government. To illustrate this, on the following figures we show an example of how two pairs of nodes which should belong to the same communities respectively, are split into four different communities with the Louvain algorithm. On one hand, we have nodes '38' and '115', both left-wing political parties that teamed back in march 2019. On the other hand, we have nodes '76', a right-wing political party, and '203', a member of this political group. After applying the Polarization Louvain algorithm, those pairs are clustered into the same communities (see Figures 13a and 13b). Let us note that mentioned images are a zoom over the whole

network, so not all the edges incident in these nodes are shown. Although it may seem that some nodes grouped in the same communities are not connected by edges (for example, nodes '76' and '203' in the image 13b) all of them are properly connected in the network.



(a) Nodes "38", "76", "115" and "203" grouped by Louvain algorithm.

(b) Nodes "38", "76", "115" and "203" grouped by Polarization Louvain algorithm.

## 7. Conclusions

In this paper we work in the definition of a polarization fuzzy measure obtained from a Polarization measure. It is a model to represent the capacity of a set of elements to argue. Then, introduce a new tool which combines the capacity represented by that polarization fuzzy measure, with the connections between elements modeled by a graph: the polarization extended fuzzy graph.

In order to handle situations in which the interest is not in the capacity of the elements to argue, but it is in their capacity to peacefully dialogue, we suggest the definition of the non-polarization fuzzy measure. Similarly as it is proposed concerning the polarization fuzzy measure, we introduce the non-polarization extended fuzzy graph, which allows the representation of the capacity to dialogue of a set of elements combined with their connections throughout a graph.

Then, we address the community detection problem in an extended context concerning the existence of several criteria to be taken into account. On the one hand, we consider the representation of the direct connections between the individuals represented by a crisp network  $G = (V, E)$ . On the other hand, we know the position of all the elements in any attitudinal axis, information not inherent to the structural representation of their connections.

From this extra-information, understood as the membership degree of each element to two extreme poles, it is defined the  $JDJ_{pol}$  polarization measure, which will be the base of characterization of a non-polarization fuzzy measure  $\mu_P$ . Then, we define the non-polarization extended fuzzy graph  $\tilde{G} = (V, E, \mu_P)$ , over which we set the basis of the community detection problem based on fuzzy measures. On this assumption, we address a real case obtained from Twitter.

The graphic representation of a network reflects the structure of a given set of nodes according to their interactions and behaviour. From this point of view, the sociological phenomenon which drive all these interactions is called homophily [60]. Conforming to the concept of homophily, a set of individuals or nodes get grouped and interact between them according to their similarities. So that, the concept of Polarization is emanated from homophily and more specifically, homogeneity [53], appearing in those scenarios where a set of nodes or individuals are split into two opposite groups. In this vein, the measurement of Polarization provides the adequate clues for community detection problems. Furthermore, the fuzzy-set theoretical approach provides the appropriate resources in order to tackle this issue from a realistic position. Adding the extra-information provided by  $JDJ_{pol}$  has

a double benefit: (1) not only allows to increase the homogeneity degree intra-community but (2) provides essential information in those cases where there are some nodes with a non-clear membership with the classical community detection algorithms. Furthermore, it is worth mentioning the importance of the aims and hypothesis of the study which should be the same for both, community detection application and Polarization measurement. Thus, the produced synergy between community detection algorithm and other measures will be an optimal solution. As a consequence, not only the integration of a given community is more realistic but the global topographic structure of a network is as well.

**Author Contributions:** For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used “conceptualization, I.G., J.A.G. and D.G.; methodology, I.G., J.A.G. and D.G.; software, I.G. and J.A.G.; validation, D.G., J.C. and R.E.; formal analysis, I.G. and J.A.G.; investigation, I.G., J.A.G.; resources, I.G., J.A.G., D.G., J.C. and R.E.; data curation, I.G. and J.A.G.; writing—original draft preparation, I.G. and J.A.; writing—review and editing, I.G., J.A.G., D.G., J.C. and R.E.; visualization, I.G. and J.A.G.; supervision, D.G., J.C. and R.E.; project administration, D.G., J.C. and R.E.; funding acquisition, D.G., J.C. and R.E.”

**Funding:** This research has been partially supported by the Government of Spain, Grant Plan Nacional de I+D+i, MTM2015-70550-P, PGC2018096509-B-I00, TIN2015-66471-P and PID2019-106254RB-I00.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

MDPI    Multidisciplinary Digital Publishing Institute  
SNA    Social Networks Analysis

1. Esteban, J.M.; Ray, D. On the measurement of polarization. *Econometrica: Journal of the Econometric Society* **1994**, *62*, 819–851.
2. Wolfson, M.C. When inequalities diverge. *The American Economic Review* **1994**, *84*, 353–358.
3. Guevara, J.A.; Gómez, D.; Robles, J.M.; Montero, J. Measuring Polarization: A Fuzzy Set Theoretical Approach. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. Springer, 2020, pp. 510–522.
4. Newman, M. Fast algorithm for detecting community structure in network. *Physical Review E* **2004**, *69*.
5. Clauset, A.; Newman, M.; Moore, C. Finding community structure in very large networks. *Physical Review E* **2004**, *70*, 066111.
6. Newman, M. Communities, modules and large-scale structure in networks. *Physics Review* **2012**, *8*, 25–31.
7. Girvan, M.; Newman, M. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **2002**, *99*, 7821–7826.
8. Blondel, V.; Guillaume, J.; Lambiotte, R.; Lefevre, E. Fast unfolding of communities in large networks. *Journal Of Statistical Mechanics-Theory And Experiment* **2008**, *10*.
9. de Blas, C.S.; Martin, J.S.; Gomez, D. Combined social networks and data envelopment analysis for ranking. *European Journal of Operational Research* **2018**, *266*, 990–999.
10. Saad, H.; Nosratinia, A. Community Detection With Side Information: Exact Recovery Under the Stochastic Block Model. *IEEE Journal of Selected Topics in Signal Processing* **2018**, *12*, 944–958.
11. Gómez, D.; González-Arangüena, E.; Manuel, C.; Owen, G.; del Pozo, M.; Tejada, J. Centrality and power in social networks: a game theoretic approach. *Mathematical Social Sciences* **2003**, *46*, 27–54.
12. Wu, J.; Chiclana, F. A social network analysis trust-consensus based approach to group-decision making problems with interval-valued fuzzy reciprocal preference relations. *Knowledge-Based Systems* **2014**, *59*, 97–107.
13. Gutiérrez, I.; Gómez, D.; Castro, J.; Espínola, R. A New Community Detection Algorithm Based on Fuzzy Measures. *Advances in Intelligent Systems and Computing series*; Kahraman, C.; Cebi, S.; Cevik Onar, S.;

- Oztaysi, B.; Tolga, A.; Sari, I., Eds. Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making. INFUS 2019, Springer, Cham, 2020, Vol. 1029, pp. 133–140.
14. Gutiérrez, I.; Gómez, D.; Castro, J.; Espínola, R. Fuzzy Measures: A solution to deal with community detection problems for networks with additional information. *Journal of Intelligent & Fuzzy Systems* **2020**, *39*, 6217–6230. doi:10.3233/JIFS-18909.
  15. Gutiérrez, I.; Gómez, D.; Castro, J.; Espínola, R. Multiple bipolar fuzzy measures: an application to community detection problems for networks with additional information. *International Journal of Computational Intelligence Systems* **2020**, *13*, 1636–1649. doi:https://doi.org/10.2991/ijcis.d.201012.001.
  16. Rosenfeld, A. Fuzzy Graphs. *Fuzzy Sets and their Applications* **1975**, pp. 77–95.
  17. Zadeh, L. Fuzzy sets. *Information and Control* **1965**, *8*, 338–353.
  18. Yaqoob, N.; Gulistan, M.; Kadry, S.; Wahab, H. Complex Intuitionistic Fuzzy Graphs with Application in Cellular Network Provider Companies. *Mathematics* **2019**, *7*, 35.
  19. Zuo, C.; Pal, A.; Dey, A. New Concepts of Picture Fuzzy Graphs with Application. *Mathematics* **2019**, *7*, 470.
  20. Shao, Z.; Kosari, S.; Rashmanlou, H.; Shoaib, M. New Concepts in Intuitionistic Fuzzy Graph with Application in Water Supplier Systems. *Mathematics* **2020**, *8*, 1241.
  21. Gómez, D.; Rodriguez, J.T.; Montero, J.; Bustince, H.; Barrenechea, E. n-Dimensional overlap functions. *Fuzzy Sets and Systems* **2016**, *287*, 57–75.
  22. Mordeson, J.; Nair, P. Fuzzy Graphs and Fuzzy Hypergraphs. *Studies in Fuzziness and Soft Computing* **2000**, *46*, 19–81.
  23. Beliakov, G. On random generation of supermodular capacities. *IEEE Transactions on Fuzzy Systems* **2020**, pp. 1–1. doi:10.1109/TFUZZ.2020.3036699.
  24. Sugeno, M. Fuzzy measures and fuzzy integrals: A survey. *Fuzzy Automata and Decision Processes* **1977**, *78*, 89–102.
  25. Newman, M. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* **2006**, *103*, 8577–8582.
  26. Newman, M.; Girvan, M. Finding and evaluating community structure in networks. *Physical Review E* **2004**, *69*.
  27. Mei, J.; He, S.; Shi, G.; Wang, Z.; Li, W. Revealing network communities through modularity maximization by a contraction-dilation method. *New Journal of Physics* **2009**, *11*, 043025.
  28. Reynal-Querol, M. Ethnic and religious conflicts, political systems and growth. PhD thesis, London School of Economics and Political Science (University of London), 2001.
  29. Apouey, B. Measuring health polarization with self-assessed health data. *Health Economics* **2007**, *16*, 20.
  30. Permanyer, I.; D'Ambrosio, C. Measuring social polarization with ordinal and categorical data. *Journal of Public Economic Theory* **2015**, *17*, 311–327.
  31. Foster, J.E.; Wolfson, M.C. Polarization and the Decline of the Middle Class: Canada and the US Mimeo. *Vanderbilt University* **1992**, *31*.
  32. Montalvo, J.G.; Reynal-Querol, M. The effect of ethnic and religious conflict on growth. *Master California* **2002**.
  33. Cheng, T.; Li, Q.; Yang, J.; Cong, G.; Li, G. Modeling of the Public Opinion Polarization Process with the Considerations of Individual Heterogeneity and Dynamic Conformity. *Mathematics* **2019**, *7*, 917.
  34. Bustince, H.; Pagola, M.; Mesiar, R.; Hullermeier, E.; Herrera, F. Grouping, overlap, and generalized bientropic functions for fuzzy modeling of pairwise comparisons. *IEEE Transactions on Fuzzy Systems* **2011**, *20*, 405–415.
  35. Grabisch, M. *k*-order additive discrete fuzzy measures and their representation. *Fuzzy Sets Systems* **1997**, *92*, 167–189.
  36. Guimera, R.; Amaral, L. Functional cartography of complex metabolic networks. *Nature* **2005**, *433*, 895–900.
  37. Flake, G.; Lawrence, S.; Giles, C.; Coetzee, F. Self-organization and identification of web communities. *Computer* **2002**, *35*, 66–70.
  38. Zou, F.; Chen, D.; Li, S.; Lu, R.; Lin, M. Community detection in complex networks: Multi-objective discrete backtracking search optimization algorithm with decomposition. *Applied Soft Computing* **2017**, *53*, 285–295.
  39. Liu, J.; Wang, J.; Liu, B. Community Detection of Multi-Layer Attributed Networks via Penalized Alternating Factorization. *Mathematics* **2020**, *8*, 239.

40. Fortunato, S. Community detection in graphs. *Physics Reports-Review Section of Physics Letters* **2010**, 486, 75–174.
41. Gupta, S.; Mittal, S.; Gupta, T.; Singhal, I.; Gupta, A.; Kumar, N. Parallel quantum-inspired evolutionary algorithms for community detection in social networks. *Applied Soft Computing* **2017**, 61, 331–353.
42. Vitali, S.; Battiston, S. The Community Structure of the Global Corporate Network. *SSNR Electronic Journal* **2013**, 8.
43. Carnivali, G.; Vieira, A.; Ziviani, A.; Esquef, P. CoVeC: Coarse-Grained Vertex Clustering for Efficient Community Detection in Sparse Complex Networks. *Information Sciences* **2020**, 522, 180–192.
44. Riolo, M.; Newman, M. Consistency of community structure in complex networks. *Physical Review E* **2020**, 101, 052306.
45. Gutiérrez, I.; Gómez, D.; Castro, J.; Espínola, R. A new community detection problem based on bipolar fuzzy measures. *Studies in Computational Intelligence*. In Press.
46. Gutiérrez, I.; Gómez, D.; Castro, J.; Espínola, R. Fuzzy Sugeno  $\lambda$ -Measures and Theirs Applications to Community Detection Problems; IEEE International Conference on Fuzzy Systems, , 2020; pp. 1–6. doi:10.1109/ISKE.2017.8258764.
47. Barroso, M.; Gutiérrez, I.; Gómez, D.; Castro, J.; Espínola, R. Group Definition Based on Flow in Community Detection. Information Processing and Management of Uncertainty in Knowledge-Based Systems; Lesot, M.J.; Vieira, S.; Reformat, M.Z.; Carvalho, J.P.; Wilbik, A.; Bouchon-Meunier, B.; Yager, R.R., Eds.; Springer International Publishing: Cham, 2020; pp. 524–538.
48. Gutiérrez, I.; Barroso, M.; Gómez, D.; Castro, J.; Espínola, R. Pattern-based clustering problem based on fuzzy measures. *Developments of Artificial Intelligence Technologies in Computation and Robotics* **2020**, 12, 412–420. doi:10.1109/ISKE.2017.8258764.
49. Shapley, L. A value for  $n$ -person games. *Contribute. Theory Games* **1953**, 2, 307–317.
50. Grabisch, M.; Nguyen, H.; Walker, E. *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference*; Kluwer Academic: Dordrecht, 1995.
51. Castro, J.; Gómez, D.; Molina, E.; Tejada, J. Improving polynomial estimation of the Shapley value by stratified random sampling with optimum allocation. *Computers & Operations Research* **2017**, 82, 108–188.
52. Castro, J.; Gómez, D.; Tejada, J. Polynomial calculation of the Shapley value based on sampling. *Computers & Operations Research* **2009**, 36, 1726–1730.
53. Robles, J.M.; Atienza, J.; Gómez, D.; Guevara, J.A. La polarización de “La Manada”. El debate público en España y los riesgos de la comunicación política digital. *Tempo Social* **2019**, 31, 193–216.
54. Kearney, M.W. rtweet: Collecting and analyzing Twitter data. *Journal of Open Source Software* **2019**, 4, 1829. R package version 0.7.0, doi:10.21105/joss.01829.
55. Wang, Z.Q.; Sun, X.; Zhang, D.X.; Li, X. An optimal SVM-based text classification algorithm. 2006 International Conference on Machine Learning and Cybernetics. IEEE, 2006, pp. 1378–1381.
56. Huang, J.; Ling, C. Using AUC and accuracy in evaluating learning algorithms **2005**. 17, 299–310.
57. Joachims, T. Text categorization with Support Vector Machines: Learning with many relevant features. Machine Learning: ECML-98; Nédellec, C.; Rouveirol, C., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 1998; pp. 137–142.
58. Park, J.; Yoon, S.; Lee, C.; Kim, J. A Simple Method for Network Visualization. *Mathematics* **2020**, 8, 1020.
59. Almende B.V.; Thieurmél, B.; Robert, T. *visNetwork: Network Visualization using 'vis.js' Library*, 2019. R package version 2.0.9.
60. McPherson, M.; Smith-Lovin, L.; Cook, J. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* **2001**, 27, 415–444.

**Algorithm 1** Polarization Louvain

---

```

1: Input:  $(A, \eta_{X_A}, \eta_{X_B}, \varphi, \phi, N)$ ;
2: Output:  $P$ ;
3: Preliminary
4:  $\mu_P \leftarrow (\eta_{X_A}, \eta_{X_B}, \varphi, \phi, N)$ ;
5:  $F_{ij} = \xi \left( Sh_i(\mu_P) - Sh_i^j(\mu_P), Sh_j(\mu_P) - Sh_j^i(\mu_P) \right)$ , for all  $i, j \in V$ ;
6:  $M \leftarrow \theta(A, F)$ ;
7:  $C_i \leftarrow \{i\}$ ,  $\forall i \in V$  (each node is an isolated community);
8:  $P \leftarrow (1, 2, \dots, n)$  (initial partition);
9: end Preliminary
10: Phase 1
11:  $(o^1, \dots, o^i, \dots, o^n) \leftarrow perm(V)$ ;
12:  $stop \leftarrow 0$ ;
13: while ( $stop == 0$ ) do
14:    $stop \leftarrow 1$ 
15:   for ( $i = 1$ ) to ( $n$ ) do
16:      $H(o^i) \leftarrow (e_1, \dots, e_h)$  (find all the neighbours of  $o^i$  in  $A$ );
17:     for ( $j = 1$ ) to ( $h$ ) do
18:       Compute  $\Delta Q_{o^i}(e_j)$  in  $M$ ;
19:     end for
20:      $j^* \leftarrow \left\{ e_\ell \mid \Delta Q_{o^i}(j^*) = \max_{\ell \in \{1, \dots, h\}} \{ \Delta Q_{o^i}(e_\ell) \} \right\}$ ;
21:     if ( $\Delta Q_{o^i}(j^*) > 0$ ) then
22:        $C_{P(o^i)} \leftarrow C_{P(o^i)} \setminus \{o^i\}$ ;
23:        $C_{P(j^*)} \leftarrow C_{P(j^*)} \cup \{o^i\}$ ;
24:        $P(o^i) \leftarrow P(j^*)$ ;
25:        $stop \leftarrow 0$ ;
26:     end if
27:   end for
28: end while
29: end Phase 1
30: Phase 2
31: Aggregate  $A^*$  from  $A$  (nodes of  $A^*$  are the communities found in Phase 1);
32: Aggregate  $M^*$  from  $M$  (nodes of  $M^*$  are the communities found in Phase 1);
33: if ( $A^* \neq A$ ) then
34:    $A \leftarrow A^*$ ;
35:    $M \leftarrow M^*$ ;
36:   Compute Phase 1 and Phase 2;
37: end if
38: end Phase 2
39: return( $P$ );

```

---