International Journal of Approximate Reasoning 74 (2016) 88–107

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

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

Daniel Gómez [a,∗], J. Tinguaro Rodríguez [b], Javier Yáñez [b,∗], Javier Montero [c,∗]

[a] Facultad de Estudios Estadísticos, Complutense University, Madrid, Avenida Puerta de Hierro s/n, 28040 Madrid, Spain
[b] Facultad de Matemáticas, Complutense University, Madrid, Plaza de las Ciencias 3, 28040 Madrid, Spain
[c] Instituto IGEO-UCM (CSIC), Facultad de Matemáticas, Complutense University, Madrid, Plaza de las Ciencias 3, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 30 September 2015
Received in revised form 9 February 2016
Accepted 18 March 2016
Available online 30 March 2016

Keywords:
Aggregation operators
Overlap functions
Grouping functions

ABSTRACT

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

1. Introduction

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

Community detection problems have been widely studied during the last decade (see e.g. [15,20]), with many applications to several disciplines. Discovering inherent communities and structures in a social network must be a main objective when we pursue a better understanding of a given network. Nevertheless, real communities in complex networks often present overlap, such that each vertex may occur in more than one community. Community detection problems with overlapping communities have been also studied in the literature (see [36]), with different purposes. On one hand, the main aim of this problem is to uncover communities allowing some key nodes to belong to more than one community. On the other

☆ This paper is part of the virtual special issue on IFSA – EUSFLAT 2015, Edited by Oscar Cordón, Luis Magdalena and José M. Alonso.
∗ Corresponding authors.
E-mail addresses: dagomez@estad.ucm.es (D. Gómez), jtrodriguez@mat.ucm.es (J. Tinguaro Rodríguez), jayage@ucm.es (J. Yáñez), monty@mat.ucm.es (J. Montero).

http://dx.doi.org/10.1016/j.ijar.2016.03.003
0888-613X/

hand, a related aim is to detect and identify those nodes (usually mentioned as *overlapping nodes*) that belong to more than one community. Overlapping nodes may play a special role in a complex network system, and how to detect them is indeed a very interesting issue. In this sense, it is important to remark that most known algorithms, such as divisive algorithms [16] or agglomerative algorithms [15], cannot detect them.

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

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

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

2. Preliminaries

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

2.1. Bivariate overlap and grouping functions

Aggregation is a basic and necessary tool for most knowledge-based systems. An *aggregation operator* [3,8–12,21] is usually defined as a real function A_n that, from n data items $x_1, \ldots, x_n$ in $[0,1]$, produces an aggregated value $A_n(x_1, \ldots, x_n)$ in $[0,1]$ [7,12]. Some desirable properties any aggregation operator should satisfy use to be imposed: for example, some boundary conditions (for all n, $A_n(0, \ldots, 0) = 0$ and $A_n(1, \ldots, 1) = 1$), monotonicity and continuity in each variable (see again [5, 10]). Other properties can be also imposed, as those studied in [6,19,25,31,32,34,35].

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

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

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

Definition 2.1.

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

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

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