Theoretical Computer Science 613 (2016) 94–114

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Min–max communities in graphs: Complexity and computational properties ☆

Miriam Di Ianni [a], Giorgio Gambosi [a,*], Gianluca Rossi [a], Paola Vocca [b]

[a] Dipartimento di Ingegneria dell'Impresa "Mario Lucertini", Università degli Studi di Roma "Tor Vergata", Roma, Italy
[b] Dipartimento di Scienze Umanistiche della Comunicazione e del Turismo, Università degli Studi della Tuscia, Viterbo, Italy

ARTICLE INFO

Article history:
Received 19 August 2014
Received in revised form 13 November 2015
Accepted 20 November 2015
Available online 26 November 2015
Communicated by M. Latapy

Keywords:
Graph theory
Community discovery
Graph partitioning

ABSTRACT

Community detection in graphs aims at identifying modules within a network and, possibly, their hierarchical organization by only using the information encoded in the graph modeling the network. Generally speaking, a community in a network is a subset of its nodes showing higher degree of interconnection with each other than to the remaining nodes. This is an informal characterization and different formal definitions of communities have been proposed in the literature, also in relation to the available information. For most such definitions, the problem of detecting a proper partition of the given network into a prefixed number of community is **NP**-hard.

In this paper, we consider the case in which a weight is associated to each edge of the graph, measuring the amount of interconnection between the corresponding pair of nodes. Under this hypothesis, we introduce and explore a new definition of community, that is, *min–max communities*, to model highly connected sets of nodes: a min–max community is a set of vertices such that the weakest (minimal) relation inside the community is stronger than the strongest (maximal) relation outside. By analyzing the structural properties induced by this definition, we prove that the problem of partitioning a complete weighted graph into $k > 0$ communities is tractable. We also show that a slight modification to this framework results into an **NP**-complete problem.

1. Introduction

The size of real networks has grown considerably in the latest years, reaching millions or even billions of nodes. The need to deal with this large number of units has produced a deep change in the way graphs are approached. In fact, graphs representing real systems are far from regular, with a coexistence of order and disorder. In the classical Erdös–Rényi random graph model the probability of having an edge between a pair of nodes is the same for all possible pairs: that is, in such graphs the distribution of edges among the nodes is highly homogeneous. Real networks are not random graphs, as they display a large amount of inhomogeneity. For example, the degree distribution is broad, with a tail that often follows a power law: therefore, many nodes with low degree coexist with some nodes with high degree. Furthermore, the distribution of edges is not only globally, but also locally inhomogeneous, with high concentrations of edges within

☆ Partially supported by Italian MIUR – PRIN 2010-11 Project n. 2010N5K7EB_002 "ARS TechnoMedia".
* Corresponding author.
E-mail address: giorgio.gambosi@uniroma2.it (G. Gambosi).

http://dx.doi.org/10.1016/j.tcs.2015.11.034
0304-3975/

special groups of nodes, and low concentrations between these groups. The dense regions identify homogeneous groups, the so-called *communities* or *clusters* [13,18,28].

The meaning of a community depends on the specific network. For example, in a social network a dense region may correspond to a real social group [4,5]; in the web graph or in a citation network a dense region may represent pages or papers on related topics [11,27]; in a protein interaction network a dense region may represent a group of proteins that have similar (mal)functions [13]; also, it is widely believed that modular structures in a biological network play an important role in biological functions [26].

The identification of communities provides a better understanding of overall properties and relations in the network. As an example, the identification of communities in biological systems may lead to a deeper comprehension of the overall structure modeled by the network [18]. Furthermore, it is expected to be relevant in the design of more efficient algorithms and techniques: for example, in the case of Online Social Networks, it may help to devise more effective business strategies and may also have direct implications on the design of the networks themselves [21,10]. A further modern scenario is that of *Opportunistic Networks* for which recent studies have shown that *social-aware* protocols provide efficient solutions for basic communication tasks [6,30,32,7].

Providing algorithmic methods to identifying communities in a network thus plays a crucial role in many fields, as pointed out in [3,8,23]: in particular, methods for community detection have been studied in several research areas, such as social networks, communication networks, biological systems [18,9].

Related literature The paper by Weiss and Jacobson [34] can be considered the first study concerning community detection. After that, a very large amount of contribution have been published: due to the width of the setting, different formal definitions of a community have been proposed; see [13] for an exhaustive survey.

In this paper, our interest is focused on communities defined in terms of what in [13] are called *local properties*, that is, properties of a specific subgraph, possibly including its immediate neighborhood, but neglecting the rest of the graph. Among local properties, a wide attention has been given to the degree of interaction of nodes towards members of the same community, compared to the degree of interaction towards the other nodes. Actually, this is what is usually considered in the most recent definitions of community. However, the first time this features has been considered is not recent and stems from social network analysis. An *LS-set* [20], or *strong community* [25], is a subgraph such that the internal degree of each node is greater than its external degree. This condition is quite strict and can be relaxed into the so-called definition of *weak community* [25] or *satisfactory partition* of a graph [14,15] or *web community* [11], for which it suffices that the external degree of the subgraph is not larger than its internal degree. The degrees of nodes inside and outside the community is also taken into account in [33] in order to define *kernel communities*, which can also be applied to identify influential users in a social network.

In [14,15] Gerber and Kobler studied the problem of partitioning a graph into two communities from a graph-theoretic point of view. After having shown that some classes of graphs are not partitionable into two communities, the authors proved the **NP**-hardness of a first generalization of this problem where nodes are weighted and a node partition into two nonempty subsets is searched such that for each node the sum of weights of its neighbors in the same subset is at least as large as the sum of weights of its neighbors in the other one. Another generalization where edges are weighted was also proved to be **NP**-hard. Variations of this problem were studied in [1,12,16,17,29]: in all such cases the problem has been proved to be **NP**-hard.

A generalization of the satisfactory partition problem was studied in [1], where each node v is required to have at least $s(v)$ neighbors in its own part, for a given function s representing the degree of satisfiability. Obviously, a satisfactory partition corresponds to $s(v) = \lceil d(v)/2 \rceil$, where $d(v)$ is the degree function. Stiebitz proved in [31] that if $s(v) \leq \lceil d(v)/2 \rceil - 1$ then such a partition always exists, and in [1] a polynomial-time algorithm is provided that finds it. In [1] it was also proved that for $\lceil d(v)/2 \rceil + 1 \leq s(v) \leq d(v) - 1$ the problem is **NP**-complete: only the complexity for $s(v) = \lceil d(v)/2 \rceil$ was left open. This gap has been closed in [2] with a proof of the **NP**-completeness of the problem; in the same paper the generalization to partitioning into $k \geq 2$ communities has also been introduced, and its hardness has been proved in all the considered cases.

Our contribution In this paper, we introduce a framework of a weighted network model and a community definition for it inspired by the one in [20,25,14,15,11,19,22,33]. Assuming the amount of similarity between nodes is provided by given edge weights, we are interested in the amount of similarity of a node with any member of its own community, compared to the one with any remaining node.

To provide some intuition about this choice, we consider a couple of examples: in a social network an edge weight may stand for the number of meetings (whatever meaning of meeting we assume) between its two ends; in the web graph, it well models the strength of the relationship of two pages linking to each other. The graph model underlying this semantic of an edge weight naturally leads to consider complete graphs, with some edge having possibly weight 0 (no interaction between its ends).

Our idea of a community is the traditional one: that is, such a structure should represent a subgraph showing a high degree of cohesion, that in our setting means high similarity between any pair of its members. More precisely, we define a *weak min–max community*, briefly *weak community*, as a set of nodes such that the minimum similarity of a node with a node inside its community is not smaller than the maximum similarity with a node outside its community. In Section 3.1 we provide a polynomial time algorithm that decides if a partition into two weak communities exists, in that case also