ARTICLE IN PRESS LAA:14402

Linear Algebra and its Applications ••• (••••) •••–•••

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

A modularity based spectral method for simultaneous community and anti-community detection

Dario Fasino [a,*], Francesco Tudisco [b]

[a] Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy
[b] Department of Mathematics, University of Padua, Padua, Italy

ARTICLE INFO

Article history:
Received 18 March 2017
Accepted 1 December 2017
Available online xxxx
Submitted by Raf Vandebril

MSC:
05C50
15A42
15B99

Keywords:
Spectral methods
Modularity matrix
Stochastic block model
Inflation product

ABSTRACT

In a graph or complex network, communities and anti-communities are node sets whose modularity attains extremely large values, positive and negative, respectively. We consider the simultaneous detection of communities and anti-communities, by looking at spectral methods based on various matrix-based definitions of the modularity of a vertex set. Invariant subspaces associated to extreme eigenvalues of these matrices provide indications on the presence of both kinds of modular structure in the network. The localization of the relevant invariant subspaces can be estimated by looking at particular matrix angles based on Frobenius inner products.

1. Introduction

This paper addresses the problem of grouping nodes of a network into communities and anti-communities, possibly emerging from a neutral background. A community is

* Corresponding author.
E-mail addresses: dario.fasino@uniud.it (D. Fasino), francesco.tudisco@math.unipd.it (F. Tudisco).

https://doi.org/10.1016/j.laa.2017.12.001

roughly defined as a set of nodes being highly connected inside and poorly connected with the rest of the graph. Conversely, an anti-community is a node set being loosely connected inside but having many external connections. Revealing these structures in data and networks is a challenging and relevant problem which has applications in many disciplines, ranging from computer science to physics and several natural and social sciences [3,6,14,16,17].

In order to address this problem from the mathematical point of view one needs a quantitative definition of what a community and an anti-community is. To this end several merit functions have been introduced in the recent literature. A very popular and fruitful idea is based on the concept of modularity, originally introduced for community detection in the statistical mechanics literature [18,19]. The modularity measure of a set of nodes $S \subset V$ in a graph $G = (V, E)$ quantifies the difference between the actual weights of edges in S with respect to the expected weight, if edges were placed at random according to a prescribed *null model*. The modularity-based criterion for community detection thus identifies a subset S as a community if its modularity measure is "large", and as an anti-community if its modularity measure is "small". The (anti-)community detection problem thus boils down to a combinatorial optimization problem, whose solution is typically approximated through a matrix-based technique which exploits the spectrum of a suitably defined *modularity matrix*.

In this work we show that dominant eigenvalues of generalized modularity matrices can be used to simultaneously look for communities and anti-communities. We propose a spectral method based on the eigenspaces associated with those eigenvalues and relate its performance to certain matrix angles. We then analyze the stochastic block model, one of the most useful generative models in community detection, to obtain indications on the average performance of the proposed method. To that goal, we characterize the dominant eigenvalues and eigenvectors of the average modularity matrix in the model. A couple of numerical experiments are included to validate the proposed computational strategy.

1.1. Notations and preliminaries

In the sequel we give a brief review of standard concepts and symbols from algebraic graph theory that we will use throughout the paper. We assume that $G = (V, E)$ is a finite, undirected graph where V and E are the vertex and edge sets, respectively. We will identify V with $\{1, \dots, n\}$. We denote adjacency of vertices i and j as $ij \in E$. We allow positive weights on both the vertex and the edge sets which we denote by $\mu : V \to \mathbb{R}_+$, $i \mapsto \mu(i)$ and $w : E \to \mathbb{R}_+$, $ij \mapsto w_{ij}$, respectively.

The symbol A denotes the adjacency matrix of G, that is, $A = (a_{ij})$ where $a_{ij} = w_{ij}$ if $ij \in E$, and $a_{ij} = 0$ otherwise. In particular, A is a symmetric, componentwise nonnegative matrix. The (generalized) degree of vertex $i \in V$ is $d_i = \sum_{j=1}^{n} w_{ij}$, and $\mathbb{1}$ denotes an all-one vector whose dimension depends on the context. With this notation, the degree vector is $d = A\mathbb{1}$.

Please cite this article in press as: D. Fasino, F. Tudisco, A modularity based spectral method for simultaneous community and anti-community detection, Linear Algebra Appl. (2018), https://doi.org/10.1016/j.laa.2017.12.001