3rd *BYMAT Conference* (2020)

Recent results and open problems in spectral algorithms for signed graphs

Bruno Ordozgoiti
Aalto University
bruno.ordozgoiti@aalto.fi

Abstract: In signed graphs, edges are labeled with either a positive or a negative sign. This small modification greatly enriches the representation capabilities of graphs. However, their spectral properties undergo significant changes, introducing new challenges in related optimization problems. In this extended abstract we discuss recent results in spectral methods for signed graph partitioning and community detection, and propose open problems arising in this context.

Resumen: En grafos con signos, las aristas se etiquetan con un signo positivo o un signo negativo. Esta pequeña modificación enriquece notablemente las capacidades de representación de los grafos. Sin embargo, sus propiedades espectrales sufren cambios significativos, lo que introduce nuevos desafíos en problemas de optimización relacionados. En este resumen extendido hablaremos de resultados recientes en métodos espectrales para la partición y detección de comunidades en grafos con signos, y propondremos problemas abiertos que surgen en este contexto.

Keywords: spectral graph theory, signed graphs.

MSC2010: 05C85.

Acknowledgements: The author thanks his co-authors in the works discussed here, namely Francesco Bonchi, Edoardo Galimberti, Aristides Gionis, Giancarlo Ruffo and Ruo-Chun Tzeng. This work was supported by the Academy of Finland project AIDA (317085) and the EC H2020RIA project "SoBigData++" (871042).

Reference: Ordozgoiti, Bruno. "Recent results and open problems in spectral algorithms for signed graphs". In: *TEMat monográficos*, 2 (2021): *Proceedings of the 3rd BYMAT Conference*, pp. 215-218. ISSN: 2660-6003. URL: https://temat.es/monograficos/article/view/vol2-p215.

1. Introduction

A signed graph can be characterized by the triple $G = (V, E, \sigma)$, with vertex set V, edge set $E \subseteq V \times V$ and signature $\sigma : E \to \{-, +\}$. We consider G to be undirected, and thus $(i, j) \in E$ if and only if $(j, i) \in E$ and $\sigma(i, j) = \sigma(j, i)$ (we omit parentheses for clarity of exposition).

We focus on the spectral analysis of signed graphs, which concerns associated matrices. Given a signed graph $G = (V, E, \sigma)$, we define the adjacency matrix $A = (a_{ij})$, $i, j \in V$, where $a_{ij} = 0$ if $(i, j) \notin E$. Otherwise, $a_{ij} = 1$ if $\sigma(i, j) = +$ and -1 if $\sigma(i, j) = -1$.

We will discuss two recent results from the computer science literature, involving the detection of subgraphs with particular characteristics. The problems, as formulated, are hard to optimize and reveal key differences with respect to their unsigned counterparts. This leads to further questions which we formulate as open problems.

2. Results in spectral signed graph partitioning

2.1. Partitioning and community detection

The first result we discuss was found by Bonchi et al. [1], and it involves a randomized algorithm to simultaneously find and partition a dense subgraph with approximation guarantees. In particular, the problem in question is formulated as follows:

Problem 1. Given a signed graph with $n \times n$ adjacency matrix A find

$$\max_{x \in \{-1,0,1\}^n \setminus \{\mathbf{0}\}} \frac{x^T A x}{x^T x}, \quad ◂$$

where $\mathbf{0}$ denotes the null vector in $\mathbb{R}^n$. This problem is akin to finding the densest subgraph, measured by average degree. However, we aim not only to find a subgraph with high average degree, but also that we can approximately partition in accordance to the edge signs; i.e., so that most edges traversing the boundary of the partition are negative. Contrary to its unsigned counterpart, this problem is NP-hard [1].

The main result of the cited work can be stated as follows:

Theorem 2. *There exists a randomized polynomial-time algorithm that outputs a vector $x \in \{-1, 0, 1\}^n \setminus \{\mathbf{0}\}$ satisfying*

$$\frac{x^T A x}{x^T x} \geq \Omega(n^{-1/2}) \lambda_1,$$

where λ_1 is the largest eigenvalue of A.

The result is tight, as there exist graphs where the gap between the optimal vector and λ_1 matches the bound [2].

2.2. Partitioning into an arbitrary number of groups

The next result we discuss is an extension of the work mentioned above —which is limited to a two-way partition of the detected subgraph— to handle an arbitrary number of subgraphs. This extension was accomplished in subsequent work by Tzeng et al. [2]. By defining the sets $E_+(G) = \{e \in E : \sigma(e) = +\}$ and $E_-(G) = \{e \in E : \sigma(e) = -\}$, we can simultaneously quantify the density and the quality of a partition using the following function, mapping collections of k disjoint vertex subsets to the reals:

$$f(S_1, \ldots, S_k) = \frac{\sum_{h \in [k]} (|E_+(S_h)| - |E_-(S_h)|) + \frac{1}{k-1} \sum_{h \neq l \in [k]} (|E_-(S_h, S_\ell)| - |E_+(S_h, S_\ell)|)}{|\bigcup_{h \in [k]} S_h|}.$$

We abuse the notation and use vertex subsets in lieu of the corresponding induced subgraphs. $E_+(S_i, S_j)$ (resp. $E_-(S_i, S_j)$) denotes the set of positive (resp. negative) edges with one endpoint in S_i and the other in S_j. $[k]$ is the set $\{1, \dots, k\}$.

We can thus formulate the problem as follows. Given a signed graph G with n vertices, the goal is to find k disjoint vertex subsets, for a given $0 \le k \le n$, attaining the following optimum:

$$\max_{S_1,\dots,S_k} f(S_1, \dots, S_k). \tag{1}$$

Some manipulations reveal that the numerator of the above problem is equivalent to the following quantity:

$$\frac{\langle A, XL_kX^T\rangle}{k-1},$$

where

- $\langle A, B\rangle$ is the Frobenius product of matrices A and B,
- $L_k = kI_k - J_k$,
- I_k is the identity matrix of order k,
- J_k is the all-ones square matrix of order k, and
- $X \in \{0,1\}^{n\times k}$ is a vertex-subset indicator matrix, so that $x_{ij} = 1$ if $i \in S_j$, $x_{ij} = 0$ otherwise.

The key insight now is that L_k has a $(k-1)$-dimensional invariant subspace, which enables the design of effective algorithms. This is because we can choose the eigenvectors so that they resemble a discrete structure. In particular, let $L_k = UDU^T$, $Y = XU$. Then, U can be chosen as follows:

$$\begin{array}{lll} (U_{:,1})^T = 1/\sqrt{k}\,[1,\dots,1], & & (U_{:,2})^T = c_1\,[k-1,-1,\dots,-1], \\ (U_{:,3})^T = c_2\,[0,k-2,-1,\dots,-1], & \dots & (U_{:,k})^T = c_{k-1}\,[0,\dots,0,1,-1]. \end{array} \tag{2}$$

A similar analysis of the cardinality of the union of the chosen sets leads to the final formulation:

$$\max_{Y\in\mathbb{R}^{n\times(k-1)}\setminus\{0\}} \frac{\mathrm{Tr}(Y^TAY)}{\mathrm{Tr}(Y^TY)} \quad \text{subject to} \quad Y_{i,j} = \begin{cases} c_j(k-j) & \text{if } i \in S_j, \\ 0 & \text{if } i \in \cup_{h=1}^{j-1} S_h \text{ or } i \notin \cup_{h\in[k]} S_h, \\ -c_j & \text{if } i \in \cup_{h=j+1}^{k} S_h. \end{cases}$$

This can be approximately optimized with approximation guarantees. In particular, the authors provide the next result [2]:

Theorem 3. *There exists a randomized polynomial-time algorithm that outputs a collection of vertex subsets $S_1, \dots, S_k$ satisfying $f(S_1,\dots,S_k) \ge \Omega\left((k\sqrt{n})^{-1}\right) OPT$,*

where OPT is the maximum of (1) over all collections of k disjoint vertex subsets.

3. Open problems

The above results highlight interesting differences arising in spectral theory when signs are introduced in graphs. Most notably, what is arguably a natural extension of the densest subgraph problem formulation becomes hard to optimize. This suggests the problem of identifying the conditions under which Problem 1 can be solved in polynomial time. We thus propose the following problem:

Problem 4. Characterize signed graphs for which Problem 1 can be solved in polynomial time. ◂

Since unsigned graphs can be seen as a special case of signed graphs, we know that the above family is non-empty. We can further extend this family by simply taking into account the spectral equivalence between unsigned and balanced graphs [3].

An additional question arises from empirical results presented in the works cited above. Despite the hardness of the problem formulations and the tightness of the given bounds, spectral algorithms usually work well in practice. This suggests the existence of a family of problem instances in which the value of the objective is not too far detached from the maximum eigenvalue. Thus, we formulate the following question:

Problem 5. Characterize signed graphs for which

$$\max_{x\in\{-1,0,1\}^n\setminus\{\mathbf{0}\}} \frac{x^T A x}{x^T x} = \frac{\lambda_1}{o\left(\sqrt{n}\right)}.$$

That is, we aim to identify the graphs that allow us to attain approximations significantly better than those described above.

References

[1] Bonchi, Francesco; Galimberti, Edoardo; Gionis, Aristides; Ordozgoiti, Bruno, and Ruffo, Giancarlo. "Discovering polarized communities in signed networks". In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 2019, pp. 961–970.

[2] Tzeng, Ruo-Chun; Ordozgoiti, Bruno, and Gionis, Aristides. "Discovering conflicting groups in signed networks". In: *Advances in Neural Information Processing Systems* 33 (2020).

[3] Zaslavsky, Thomas. "Matrices in the Theory of Signed Simple Graphs". In: *arXiv e-prints* (2013). arXiv: 1303.3083 [math.CO].