Discrete Applied Mathematics 217 (2017) 276–285

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

# Frustration and isoperimetric inequalities for signed graphs

Florian Martin

*Philip Morris International R&D, Quai Jeanrenaud 5, 2000 Neuchatel, Switzerland*

ARTICLE INFO

*Article history:*
Received 2 April 2015
Received in revised form 29 August 2016
Accepted 4 September 2016
Available online 9 November 2016

*Keywords:*
Spectral graph theory
Signed graphs
Monotone systems
Frustration index
Signed Laplacian
Balanced graphs

ABSTRACT

Let $G_\sigma = (V, E, \sigma)$ be a connected signed graph. Using the equivalence between signed graphs and 2-lifts of graphs, we show that the frustration index of $G_\sigma$ is bounded from below and above by expressions involving another graph invariant, the smallest eigenvalue of the (signed) Laplacian of $G_\sigma$. From the proof, stricter bounds are derived. Additionally, we show that the frustration index is the solution to a $l^1$-norm optimization problem over the 2-lift of the signed graph. This leads to a practical implementation to compute the frustration index. Also, leveraging the 2-lifts representation of signed graphs, a straightforward proof of Harary's theorem on balanced graphs is derived. Finally, real world examples are considered.



## 1. Introduction

A signed graph is defined as a graph $G_\sigma = (V, E, \sigma)$ for which a sign function $\sigma : E \to \{-1, +1\}$ is defined over the edges. Such graphs appear naturally in fields such as sociology or systems biology. In sociology, it may describe relationships between individuals, whereby the relations can be "friendly" and "unfriendly". In systems biology, those graphs, called biological networks, describe activation or inhibition between molecules, typically enzymes, proteins or genes. Unless specifically stated, all the graphs considered in the following are assumed to be connected and undirected.

Balance in a signed graph is characterized by the property that every path between two nodes have the same sign (the sign of a path is the product of its edge signs). Equivalently, a graph is balanced if and only if every cycle is positive. In sociology, it is thought [7], that relationship graphs tend to be balanced (individuals tend to form complementary alliances).

In systems biology, linearization of dynamical systems between molecules around a given state leads to signed graphs: assume that a dynamical system is described by a set of $n$-differential equations.

$$\frac{\partial x}{\partial t} = f(x(t)).$$

When linearized around a state $x_0$, the signs of the Jacobian of the system define a signed graph. Monotone dynamical systems are systems whose behaviors are simple (as opposed to oscillatory or chaotic solutions). It has been shown [6] that a dynamical system is monotone if and only if the underlying undirected signed graph is balanced.

The minimum number of edges to delete in a sign graph $G_\sigma = (V, E, \sigma)$ to make the resulting graph balanced is called the *frustration index* [10] and will be denoted by $\mathcal{F}(G_\sigma)$. A signed graph has therefore a zero frustration index if and only if it is balanced. This index indicates how far a graph is from being balanced and has attracted much attention through sociology [11,7] and systems biology [16,15,23,22]. Algorithms and inequalities were established for estimating the frustration index [16,15,14,27,8,29]. Moreover, the size of modern signed graphs (hundreds or thousands of nodes and edges) makes the finding of a minimal set of edges whose deletion leads to a balanced graph challenging (this task is known to be a *NP*-hard problem).

*E-mail address:* florian.martin@pmi.com.

http://dx.doi.org/10.1016/j.dam.2016.09.015

Simple bounds have been established. For example, if the graph has $n$ nodes and $m$ edges, then $m - n + 1$ is an upper bound for the frustration index (it follows from the deletion of all the edges, not belonging to a given spanning tree). Another upper bound is given by $(m - \sqrt{m})/2$ (see e.g., [16]).

The signed adjacency matrix of $G_\sigma$, $A(G_\sigma)$, is defined as an operator on $l^2(V)$ by $A_{xy} = \sigma(x, y)$ if $x$ and $y$ are adjacent and 0 else. The signed Laplacian of $G_\sigma$ is defined as $L(G_\sigma) = D(G_\sigma) - A(G_\sigma)$ where $D(G_\sigma)$ is the diagonal matrix whose entries are the degrees of the vertices in $(V, E)$ (here, the degree in a signed graph is defined as the degree in the classical sense for the underlying unsigned graph). If the sign function is constant and positive, $L(G_\sigma)$ is the combinatorial Laplacian of the undirected unsigned graph $(V, E)$. Also, the resulting graph will be denoted by $G_+$; similarly when the sign function is constant and negative, the notation $G_-$ will be used. Finally, the spectrum of $L(G_\sigma)$, $Sp(L(G_\sigma))$, will be denoted by $0 \le \mu_1 \le \cdots \le \mu_n$ and the spectrum of an unsigned graph $(V, E)$ by $0 = \lambda_0 \le \lambda_1 \le \cdots \le \lambda_{n-1}$, where $n = |V|$ ($\lambda_0 = 0$ as the constant functions in $l^2(V)$ are in the kernel of $L(G)$).

In spectral graph theory, the combinatorial Laplacian is leveraged to characterize graph properties through spectral properties of this matrix [5]. As a simple example, the number of connected components of a graph $(V, E)$ is equal to the multiplicity of $\lambda_0$. Another example is the relationship between the Cheeger constant of a graph and its first non-vanishing eigenvalue $\lambda_1$, called Cheeger inequalities [5]. Expander graph is another graph property linked to the spectrum of the Laplacian.

By considering the signed Laplacian, attempts to establish spectral bound on the frustration index or related graph invariants have been proposed [9,2,13,12,17,1]. However, no direct spectral upper bound has been derived for the frustration index. Lift of graphs have been considered in connection with graph spectral gap [3] to construct expander graphs.

In [19], cause-and-effect biological network models are used to quantify biological network response to a treatment in a cell system, by using gene expression experimental data. This approach requires the signed graph underlying the network to be balanced. In a subsequent work [25], some unbalanced networks were manually curated by biological experts to lead to balance by deleting some edges. However, complex networks such as the Oxidative Stress and Cell cycle network could not be curated manually [24], which may prevent the use of the methodology in [19]. Therefore, finding the maximal balanced subgraph by deleting edges is of particular interest.

The manuscript is organized as follows: In Section 2, we recall the definition and some basic facts about graph 2-lifts. In Section 3, we use those notions to derive straightforward proofs of some known theorems on graph balance. We then establish a spectral upper and lower bound for the frustration index. Leveraging the proof of the main theorem of the section, we show that the frustration index can be formulated as an optimization problem over $l^1$-functions. Finally, the frustration for real-life networks is analyzed in Section 4.

## 2. 2-lifts of graphs

In this section we recall the definition and some basic facts about 2-lift (or lift for short) of graphs. As a convention, we assume that there are no self-loops in the graphs considered and that graphs are connected.

**Definition 2.1.** Let $G = (V, E)$ be a graph. A 2-*lift* of $G$ is a graph $\hat{G} = (\hat{V}, \hat{E})$ whose vertices are given by $\hat{V} = V_+ \amalg V_-$, $V_+ \cong V_- \cong V$ and where each edge $(x, y)$ of $G$ is lifted to two edges of $\hat{G}$: either $\{(x_+, y_+), (x_-, y_-)\}$ or $\{(x_-, y_+), (x_+, y_-)\}$. The natural surjective homomorphism between $\hat{G} \to G$ is called the *lift-map*.

The adjacency matrix of $\hat{G}$, $\hat{A}$, has a block structure

$$\begin{pmatrix} A_1 & A_2 \\ A_2 & A_1 \end{pmatrix}$$

where $A_1$ contains the edges within $V_+$, and $V_-$ and $A_2$ the edges between $V_+$ and $V_-$. It is straightforward to note that the adjacency matrix of $G$, $A(G)$ is given by $A_1 + A_2$. There is an obvious equivalence between the set of 2-lifts of a graph and the set of all signings of that graph. If $\sigma$ is the signing corresponding to the 2-lift, then the signed adjacency matrix of $G_\sigma$, $A(G_\sigma) = A_1 - A_2$. The lift associated to a specific signing $\sigma$ will be denoted by $\hat{G}_\sigma$ or simply $\hat{G}$ when the context is clear.

**Lemma 2.2** ([18]). *The spectrum of the lift $\hat{G}$ is the union (with multiplicities) of the spectrum of $G_\sigma$ and $G$. Moreover, the eigenvectors of $\hat{G}$ are of the two following forms: $\binom{f}{f}$ where $f$ is an eigenvector of $G$ or $\binom{f}{-f}$ where $f$ is an eigenvector of $G_\sigma$.*

**Proof.** Let $f$ be an eigenvector of $G$ with eigenvalue $\lambda$. Then $\binom{f}{f}$ is an eigenvector of $\hat{A}$ with eigenvalue $\lambda$. Alternatively if $f$ is an eigenvector of $A(G_\sigma)$ with eigenvalue $\mu$, then $\binom{f}{-f}$ is an eigenvector of $\hat{A}$ of eigenvalue $\mu$. As $\hat{A}$ is twice the size of either $A(G)$ or $A(G_\sigma)$, summing the multiplicities of $A(G)$ and $A(G_\sigma)$ shows that every eigenvector of $\hat{A}$ is of this form. ∎

The lemma above extends directly to the Laplacian matrices associated to $G$, $G_\sigma$ and $\hat{G}$ using the fact that

$$D(G) = D(G_\sigma), \qquad D(\hat{G}) = \begin{pmatrix} D(G) & 0 \\ 0 & D(G) \end{pmatrix} \quad \text{and} \quad L(\hat{G}) = \begin{pmatrix} D(G) - A_1 & A_2 \\ A_2 & D(G) - A_1 \end{pmatrix}.$$