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_σ is bounded from below and above by expressions involving another graph invariant, the smallest eigenvalue of the (signed) Laplacian of G_σ. 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_σ, $A(G_\sigma)$, 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_σ 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 $0 \leq \mu_1 \leq \cdots \leq \mu_n$ and the spectrum of an unsigned graph (V, E) by $0 = \lambda_0 \leq \lambda_1 \leq \cdots \leq \lambda_{n-1}$, 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 λ_0. Another example is the relationship between the Cheeger constant of a graph and its first non-vanishing eigenvalue λ_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 σ is the signing corresponding to the 2-lift, then the signed adjacency matrix of G_σ, $A(G_\sigma) = A_1 - A_2$. The lift associated to a specific signing σ 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_σ 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_σ.*

Proof. Let f be an eigenvector of G with eigenvalue λ. Then $\binom{f}{f}$ is an eigenvector of $\hat{A}$ with eigenvalue λ. Alternatively if f is an eigenvector of $A(G_\sigma)$ with eigenvalue μ, then $\binom{f}{-f}$ is an eigenvector of $\hat{A}$ of eigenvalue μ. 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_σ and $\hat{G}$ using the fact that

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

A function in $l^2(\hat{V})$ of the form $\begin{pmatrix} f \\ -f \end{pmatrix}$ with $f \in l^2(V)$, will be called *anti-symmetric*. Finally, we immediately derive from those considerations the following lemma:

Lemma 2.3.

$$\mu_1(G_\sigma) = \inf_{\substack{g\in l^2(V) \\ g\neq 0}} \frac{\langle L(G_\sigma)g|g\rangle}{\|g\|^2} = \inf_{\substack{f\in l^2(\hat{V}) \\ f\neq 0 \\ f \text{ anti-symmetric}}} \frac{\langle L(\hat{G}_\sigma)f|f\rangle}{\langle f|f\rangle}.$$

Observe that on one hand, $\lambda_0(\hat{G}_\sigma)$ corresponds to $\lambda_0(G_+)$ and on the other hand, $\lambda_1(\hat{G}_\sigma)$, the second eigenvalue (accounting for multiplicities) of the Laplacian of $\hat{G}_\sigma$, is either equal to $\lambda_1(G)$ or to $\mu_1(G_\sigma)$.

3. Main results

This section is devoted to the use of 2-lift of graph to study properties of a signed graph G_σ and the frustration index of such a graph.

3.1. Results based on 2-lifts

We leverage the bijection between the set of 2-lifts of a graph and all possible signings of that graph to prove some known facts on signed graphs. The next lemma relates the balanced-property of a signed graph to a topological property of the associated 2-lift:

Lemma 3.1. *Let $G_\sigma = (V, E, \sigma)$ be a connected signed graph and let $\hat{G}_\sigma$ be the associated lift. Then G_σ is balanced if and only if $\hat{G}_\sigma$ has two connected components.*

Proof. If $\hat{G}_\sigma$ is connected, then, for any $x \in V$ there exists a path from x_+ to x_-. This path involves necessarily an odd number of crossing edges from V_+ to V_- which corresponds to a negative cycle in G_σ. Conversely, if $\hat{G}_\sigma$ has two connected components, each component should project on V by the lift map, which implies that a connected component cannot contain x_+ and x_- simultaneously. This shows that every cycle in G_σ must be positive. ∎

This directly implies combining the last two lemmas that G_σ is balanced if and only if $0 \in Sp(L(G_\sigma))$, which provides thereof a short proof of this well known fact, whose proof usually involves the matrix-tree theorem for signed graphs [28].

The next result, the proof of Harary's theorem on balanced graphs [10] becomes straightforward when using the 2-lift formalism.

Theorem 3.2 (*Harary [10]*)**.** *Let $G_\sigma = (V, E, \sigma)$ be a connected signed graph. G_σ is balanced if and only if $\hat{V}$ can be partitioned in two subsets, V_1, V_2 such that all the edges within V_i's are positive and all others between V_1 and V_2 are negative.*

Proof. Let $\pi : \hat{G}_\sigma \to G$ be the covering map. From the previous lemma, G_σ is balanced if and only if $\hat{G}$ has two connected components whose vertex sets are denoted by C_1, C_2. Let $\hat{V}_1 = C_1 \cap V_+ \amalg C_2 \cap V_-$ and $\hat{V}_2 = C_1 \cap V_- \amalg C_2 \cap V_+$. Then $\hat{V}_1$ and $\hat{V}_2$ are disjoint. Then $V_i = \pi(\hat{V}_i)$ defines a partition of V. By definition of V_1, V_2, all the edges between V_1 and V_2 are negative (i.e., cross the two layers) and all the ones within V_i's are positive (i.e., are within the same layer). ∎

Two signings, σ_1, σ_2 of the graph $G = (V, E)$, are said to be *switching equivalent* if there exists a map $\theta : V \to \{1, -1\}$ such that $\sigma_2(u, v) = \theta(u)\sigma_1(u, v)\theta(v), \forall u \sim v \in E$. This notion translates to an operation of the lifted graphs which is given in the following lemma.

Lemma 3.3. *Two signings σ_1, σ_2, are equivalent if and only if there exist a sequence of permutation matrices P_i, $i = 1, \ldots, k$, each permuting vertices from V_+ and V_-, such that $P^T A(\hat{G_{\sigma_2}})P = A(\hat{G_{\sigma_2}})$ where $P = P_1 \cdots P_k$.*

The proof is immediate and will be omitted. For a subset of vertices S in a graph we denote by ∂S the boundary of S, which is defined as the set of vertices not in S adjacent to at least one vertex in S. The usage of 2-lifts for solving the frustration index problem is justified in the following lemma that is an immediate consequence of the definitions.

Lemma 3.4. *Let $G_\sigma = (V, E, \sigma)$ be a connected signed graph and let $\hat{G}_\sigma$ be the associated 2-lift. A partition of $\hat{V}$ is said to be admissible in the lift if $|\hat{V}_1| = |\hat{V}_2| = n$ and $V = \pi(\hat{V}_i)$ for each partition set. We have*

$$\frac{2}{n}\mathcal{F}(G_\sigma) = \inf \frac{|\partial \hat{V}_1|}{|\hat{V}_1|},$$

where infimum is taken over all admissible partitions $\hat{V}_1 \amalg \hat{V}_2$. By symmetry of the lift $\hat{V}_2$ can be equally used in the above equation.

3.2. The frustration index and $\mu_1(G_\sigma)$

The following theorem establishes a link between the frustration index and the first eigenvalue of the signed Laplacian matrix, $\mu_1(G_\sigma)$.

Theorem 3.5. *Let $\mathcal{F}(G_\sigma)$ be the frustration index of G_σ and $\mu_1(G_\sigma)$ be the first eigenvalue of $L(G_\sigma)$. Then the following inequalities hold:*

$$\frac{n}{4}\cdot\mu_1(G_\sigma)\leq\mathcal{F}(G_\sigma)\leq\frac{n}{\sqrt{2}}\cdot\sqrt{\mu_1(G_\sigma)(2\Delta-\mu_1(G_\sigma))}$$

where Δ is the maximum degree in G_σ.

Proof. Let C be a minimal cut that makes the signed graph balanced and let $\hat{C}$ be its lift and $V_1\amalg V_2$ the associated partition of $\hat{V}$. Let f be the unit norm function (as $|V_1|=|V_2|=n$) defined by:

$$f(x)=\begin{cases}\dfrac{1}{\sqrt{2n}} & \text{if } x\in V_1\\ -\dfrac{1}{\sqrt{2n}} & \text{if } x\in V_2.\end{cases}$$

Using Lemma 2.3:

$$\mu_1(G_\sigma)=\inf_{\substack{f\neq 0\\ f\,\text{anti-symmetric}}}\frac{\langle L(\hat{G})f|f\rangle}{\langle f|f\rangle}=\inf_{\substack{f\neq 0\\ f\,\text{anti-symmetric}}}\frac{\sum_{x\sim y\in\hat{E}}(f(x)-f(y))^2}{\|f\|^2},$$

we have

$$\mu_1(G_\sigma)\leq\sum_{x\in V_1,y\in V_2}(f(x)-f(y))^2=\frac{2}{n}\cdot|\hat{C}|=\frac{4}{n}\cdot|C|.$$

For the second inequality, we follow the strategy of the proof in [20] and make use of the symmetry of the lift. Let us denote $\mu_1(G_\sigma)$ by μ_1, $L(\hat{G}_\sigma)$ by $\hat{L}$ and $D(\hat{G}_\sigma)$ by $\hat{D}$.

Let g be an eigenfunction for μ_1 on $\hat{G}$ (which is anti-symmetric). Let $V^+=\{x\in\hat{V}|g(x)\geq 0\}$, $V^-=\{x\in\hat{V}|g(x)<0\}$ and let g^+, g^- be defined by

$$g^+(x)=\begin{cases}g(x) & \text{if } x\in V^+\\ 0 & \text{else}\end{cases}\quad\text{and}\quad g^-(x)=\begin{cases}g(x) & \text{if } x\in V^-\\ 0 & \text{else.}\end{cases}$$

Using the Rayleigh quotient, we have:

$$\begin{aligned}\mu_1\|g^++g^-\|^2&=\langle g^++g^-|\hat{L}(g^++g^-)\rangle\\&=\langle g^+|\hat{L}g^+\rangle+\langle g^-|\hat{L}g^-\rangle+2\langle g^+|\hat{L}g^-\rangle.\end{aligned}\tag{1}$$

If $\hat{L}^-$ denotes the signed Laplacian of $\hat{G}_-$ where $-$ denotes the negative constant sign function on G, then $\hat{L}^-+\hat{L}=2\hat{D}$ which implies that for any $h\in l^2(\hat{V})$, $\langle h|\hat{L}^-h\rangle-\langle(2\hat{D}-\mu_1)h|h\rangle=\mu_1\langle h|h\rangle-\langle h|\hat{L}h\rangle$. Therefore, using this equality, the orthogonality between g^+ and g^- and, as $\hat{D}$ is diagonal, that $\langle\hat{D}g^-|g^+\rangle=0$ by definition of g^+ and g^-, we get:

$$\begin{aligned}&\langle g^+|\hat{L}^-g^+\rangle+\langle g^-|\hat{L}^-g^-\rangle-(2\Delta-\mu_1)\|g^++g^-\|^2\\&\leq\langle g^+|\hat{L}^-g^+\rangle+\langle g^-|\hat{L}^-g^-\rangle-\langle(2\hat{D}-\mu_1)(g^++g^-)|g^++g^-\rangle\\&=\langle g^+|\hat{L}^-g^+\rangle-\langle(2\hat{D}-\mu_1)g^+|g^+\rangle+\langle g^-|\hat{L}^-g^-\rangle-\langle(2\hat{D}-\mu_1)g^-|g^-\rangle-2\langle(2\hat{D}-\mu_1)g^-|g^+\rangle\\&=\mu_1\|g^++g^-\|^2-\langle g^+|\hat{L}g^+\rangle-\langle g^-|\hat{L}g^-\rangle\\&=2\langle g^+|\hat{L}g^-\rangle.\end{aligned}$$

which implies that

$$(2\Delta-\mu_1)\|g^++g^-\|^2\geq\langle g^+|\hat{L}^-g^+\rangle+\langle g^-|\hat{L}^-g^-\rangle-2\langle g^+|\hat{L}g^-\rangle.\tag{2}$$

Multiplying Eqs. (1) and (2), setting $\gamma = 2\langle g^+|\hat{L}g^-\rangle$, we get as $\hat{L}$ and $\hat{L}^-$ are positive definite:

$$\begin{aligned}\mu_1(2\Delta - \mu_1)\|g^+ + g^-\|^4 \geq\ & \langle g^+|\hat{L}g^+\rangle\langle g^+|\hat{L}^-g^+\rangle + \langle g^-|\hat{L}g^-\rangle\langle g^-|\hat{L}^-g^-\rangle \\ & + \langle g^+|\hat{L}g^+\rangle\langle g^-|\hat{L}^-g^-\rangle + \langle g^-|\hat{L}g^-\rangle\langle g^+|\hat{L}^-g^+\rangle \\ & + \gamma(-\langle g^+|\hat{L}g^+\rangle + \langle g^+|\hat{L}^-g^+\rangle - \langle g^-|\hat{L}g^-\rangle + \langle g^-|\hat{L}^-g^-\rangle - \gamma) \\ \geq\ & \langle g^+|\hat{L}g^+\rangle\langle g^+|\hat{L}^-g^+\rangle + \langle g^-|\hat{L}g^-\rangle\langle g^-|\hat{L}^-g^-\rangle \\ & + \gamma(\langle g^+|\hat{L}^-g^+\rangle + \langle g^-|\hat{L}^-g^-\rangle - \langle g^-|\hat{L}g^-\rangle - \langle g^+|\hat{L}g^+\rangle - \gamma).\end{aligned}$$

As $\hat{L}^- = 2\hat{D} - \hat{L}$,

$$\begin{aligned}\langle g^+|\hat{L}^-g^+\rangle + \langle g^-|\hat{L}^-g^-\rangle - \langle g^-|\hat{L}g^-\rangle - \langle g^+|\hat{L}g^+\rangle - \gamma &= \langle 2\hat{D}g^+|g^+\rangle + \langle 2\hat{D}g^-|g^-\rangle - \gamma - 2\langle \hat{L}g^+|g^+\rangle - 2\langle \hat{L}g^-|g^-\rangle \\ &= \langle 2\hat{D}g^+|g^+\rangle + \langle 2\hat{D}g^-|g^-\rangle - \mu_1\langle g^+|g^+\rangle - \mu_1\langle g^-|g^-\rangle + \gamma \\ &= \langle (2\hat{D} - \mu_1)g^+|g^+\rangle + \langle (2\hat{D} - \mu_1)g^-|g^-\rangle + \gamma.\end{aligned}$$

As γ is positive, the latest expression is positive if and only if $\mu_1 \leq \min_{x\in\hat{V}} \deg(x)$. Let $\{x_+, x_-\}$ be the fiber of a vertex $x \in V$ in $\hat{V}$ and let us a define an anti-symmetric function $h \in l^2(\hat{V})$ as

$$h_x(z) = \begin{cases} \dfrac{1}{\sqrt{2}} & \text{if } z = x_+ \\ -\dfrac{1}{\sqrt{2}} & \text{if } z = x_- \\ 0 & \text{else} \end{cases}$$

Using the Lemma 2.3, and the equality $\langle L(\hat{G})h|h\rangle = \deg(x_+) + \deg(x_-)$, we conclude that $2\hat{D} - \mu_1$ is positive definite.

Therefore as by definition γ is also ≥ 0:

$$\mu_1(2\Delta - \mu_1)\|g^+ + g^-\|^4 \geq \langle g^+|\hat{L}g^+\rangle\langle g^+|\hat{L}^-g^+\rangle + \langle g^-|\hat{L}g^-\rangle\langle g^-|\hat{L}^-g^-\rangle$$

Let $\alpha_+ = \sum_{x\sim y} |g^+(x)^2 - g^+(y)^2|$ and $\alpha_- = \sum_{x\sim y} |g^-(x)^2 - g^-(y)^2|$. By Cauchy–Schwarz inequality, we obtain:

$$\begin{aligned}\alpha_+^2 + \alpha_-^2 &\leq \langle g^+|\hat{L}g^+\rangle \cdot \langle g^+|\hat{L}^-g^+\rangle + \langle g^-|\hat{L}g^-\rangle \cdot \langle g^-|\hat{L}^-g^-\rangle \\ &\leq \mu_1(2\Delta - \mu_1)\|g^+ + g^-\|^4.\end{aligned} \tag{3}$$

It follows from the anti-symmetry of g, that the set of distinct values taken by g^+ is the same as the set of values taken by g^- up to the sign. Let $-\xi_m < \cdots < -\xi_1 < 0 = \xi_0 < \xi_1 < \cdots < \xi_m$ be the values taken by g. In particular, by symmetry of the lift, $\alpha_+ = \alpha_-$. Let $V_i^+ = \{x \in \hat{V}|g(x) \geq \xi_i\}$ and $V_i^- = \{x \in \hat{V}|0 > g(x) > -\xi_i\}$.

Therefore,

$$\partial V_i^+ = \{y \in \hat{V}|\exists x \sim y \text{ s.t. } g(x) \geq \xi_i > g(y)\}$$

and

$$\partial V_i^- = \{y \in \hat{V}|\exists x \sim y \text{ s.t. } 0 > g(x) > -\xi_i \text{ and } g(y) \leq -\xi_i \text{ or } g(y) \geq 0\}.$$

On one hand, we have:

$$\begin{aligned}\alpha_+ &= \sum_{i=1}^m \sum_{\substack{x\sim y \\ g^+(y)=\xi_j<g^+(x)=\xi_i}} (g^+(x)^2 - g^+(y)^2) + \sum_{i=1}^m \sum_{\substack{x\sim y \\ g^+(y)=0<g^+(x)=\xi_i}} (g^+(x)^2 - g^+(y)^2) \\ &= \sum_{i=1}^m \sum_{\substack{x\sim y \\ g^+(y)=\xi_j<g^+(x)=\xi_i}} (\xi_i^2 - \xi_j^2) + \sum_{i=1}^m \sum_{\substack{x\sim y \\ g^+(y)=0<g^+(x)=\xi_i}} \xi_i^2 \\ &\geq \sum_{i=1}^m \sum_{\substack{x\sim y \\ g^+(y)=\xi_j<g^+(x)=\xi_i}} (\xi_i^2 - \xi_{i-1}^2) + \sum_{i=1}^m \sum_{\substack{x\sim y \\ g^+(y)=0<g^+(x)=\xi_i}} (\xi_i^2 - \xi_{i-1}^2) \\ &= \sum_{i=1}^m |\partial V_i^+|(\xi_i^2 - \xi_{i-1}^2)\end{aligned}$$

and on the other hand:

$$\begin{aligned}
\alpha_- &= \sum_{i=1}^{m} \sum_{\substack{x\sim y \\ g^-(x)=-\xi_i<g^-(y)=-\xi_j}} (g^-(y)^2 - g^-(x)^2) + \sum_{i=1}^{m} \sum_{\substack{x\sim y \\ g^-(x)=-\xi_i<0\leq g(y)=\xi_j}} g^-(x)^2 \\
&= \sum_{i=1}^{m} \sum_{\substack{x\sim y \\ g^-(x)=-\xi_i<g^-(y)=-\xi_j}} (\xi_i^2 - \xi_j^2) + \sum_{i=1}^{m} \sum_{\substack{x\sim y \\ g^-(x)=-\xi_i<0\leq g(y)=\xi_j}} \xi_i^2 \\
&\geq \sum_{i=1}^{m} \sum_{\substack{x\sim y \\ g^-(x)=-\xi_i<g^-(y)=-\xi_j}} (\xi_i^2 - \xi_{i-1}^2) + \sum_{i=1}^{m} \sum_{\substack{x\sim y \\ g^-(x)=-\xi_i<0\leq g(y)=\xi_j}} (\xi_i^2 - \xi_{i-1}^2) \\
&= \sum_{i=1}^{m} |\partial V_i^-|(\xi_i^2 - \xi_{i-1}^2).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\alpha_+ + \alpha_- &\geq \sum_{i=1}^{m} (|\partial V_i^+| + |\partial V_i^-|)(\xi_i^2 - \xi_{i-1}^2) \\
&\geq \frac{2\mathcal{F}(G_\sigma)}{n} \sum_{i=1}^{m} |V_i^+ \cup V_i^-|(\xi_i^2 - \xi_{i-1}^2) \\
&= \frac{2\mathcal{F}(G_\sigma)}{n} \sum_{i=0}^{m} \xi_i^2 \left(|V_i^+ \cup V_i^-| - |V_{i+1}^+ \cup V_{i+1}^-|\right) \\
&= \frac{2\mathcal{F}(G_\sigma)}{n} \sum_{x\in\hat{V}} (g^+(x)^2 + g^-(x)^2) \qquad (4)
\end{aligned}$$

where $\xi_0 = 0$, $|V_{m+1}^+| = 0$ and $V_{m+1} = \{x \in \hat{V} | g(x) = -\xi_m\}$. It is important to note that the partition of $\hat{G}$ induced by $V_i^+ \cup V_i^-$ (and its complement) is admissible in the sense of Lemma 3.4. Using (3) and (4) and the fact that $\alpha_+ = \alpha_-$, we finally obtain:

$$\mathcal{F}(G_\sigma) \leq \frac{n}{\sqrt{2}} \sqrt{\mu_1(2\Delta - \mu_1)}$$

Note that the theorem is trivial for G_+. The result of Theorem 3.5 has to be compared with the work in [12] where another graph invariant, $\psi(G_\sigma) \doteq \min_{\emptyset\neq S\subseteq V} \frac{\mathcal{F}(S)+|\partial S|}{|S|}$, has been bounded below by $\frac{\mu_1(G_\sigma)}{4}$ and above by $\sqrt{\mu_1(G_\sigma)(2\Delta - \mu_1(G_\sigma))}$. It is clear from the definitions that $\psi(G_\sigma) \leq \mathcal{F}(G_\sigma)/n$. Therefore, using the upper bound of Theorem 3.5 leads to a bound of $\psi(G_\sigma)$ improved by a factor $1/\sqrt{2}$.

Corollary 3.6. *Let f be an anti-symmetric eigenfunction in $l^2(\hat{V})$ corresponding to μ_1. If $V_+ = \{x \in \hat{V} | f(x) > 0\}$, then the following inequalities hold*

$$\frac{n}{2} \sum_{x\sim y\in\partial V_+} (f(x) - f(y))^2 \leq \mathcal{F}(G_\sigma) \leq |\partial V_+|$$

where f is the eigenfunction associated with the first (non-zero) eigenvalue of $\hat{G}$.

Proof. By symmetry $\sum_{x\sim y\in\partial V_+} (f(x) - f(y))^2 = \sum_{x\sim y\in\partial V_-} (f(x) - f(y))^2$ and each of those terms is bounded by $\lambda_1(G_\sigma)$ by definition of f. ■

Following Corollary 3.6, we see that considering the admissible partition induced by $\mathcal{F}_{l^2}(G_\sigma) \doteq |\partial V_+|$ is a (computationally) fast approximation to the frustration index and provide a sub-optimal set of edges for solving the frustration problem.

3.3. A l^1-norm minimization formulation for the frustration index

In this section, we obtain an equality relating the frustration index to a convex optimization problem. Recall that a function $f \in l^2(\hat{V})$ is said to be *anti-symmetric* if there exists an admissible partition $V_1 \amalg V_2$ such that the restrictions to the subset V_i (denoted by $f|_{V_i}$) satisfy $f|_{V_1} = -f|_{V_2}$.

Theorem 3.7. *Let $\mathcal{F}(G_\sigma)$ be the frustration index of G_σ. The following equality holds:*

$$\mathcal{F}(G_\sigma) = \frac{n}{2} \min_{f \neq 0 \in l^2(\hat{V}),\, f \text{ anti-symmetric}} \frac{\sum_{x \sim y} |f(x) - f(y)|}{\sum_x |f(x)|}.$$

Proof. Let $f \in l^2(\hat{V})$ that is anti-symmetric ($f(x_+) = -f(x_-)$, where x_+, x_- is the fiber of $x \in V$). Let $c \in \mathbb{R}^+$. Following the same strategy of the proof of Theorem 3.5, let $-\xi_m < \cdots < -\xi_1 < 0 = \xi_0 < \xi_1 < \cdots < \xi_m$ be the values taken by f. Let f^+ and f^- defined as in Theorem 3.5 and $V_i = \{x | f(x) \geq \xi_i \text{ or } 0 > f(x) > -\xi_i\}$. Then following the arguments of the previous proof:

$$\begin{aligned}
\sum_{x \sim y} |f(x) - f(y)| &\geq \frac{2\mathcal{F}(G_\sigma)}{n} \sum_{i=1}^{m} |V_i|(\xi_i - \xi_{i-1}) \\
&= \frac{2\mathcal{F}(G_\sigma)}{n} \sum_{i=0}^{m} \xi_i \,(|V_i| - |V_{i+1}|) \\
&= \frac{2\mathcal{F}(G_\sigma)}{n} \sum_{x \in \hat{V}} (f^+(x) - f^-(x)) = \sum_{x \in \hat{V}} |f(x)|. \qquad (5)
\end{aligned}$$

This shows $\mathcal{F}(G_\sigma) \leq \frac{n}{2} \min_{f \neq 0} \frac{\sum_{x \sim y} |f(x) - f(y)|}{\sum_x |f(x)|}$.

The equality follows from considering a minimum admissible partition $V_1 \amalg V_2$ of the lift. By defining the unit l^1-norm function (as $|V_1| = |V_2| = n$):

$$f(x) = \begin{cases} \dfrac{1}{2} & \text{if } x \in V_1 \\ -\dfrac{1}{2} & \text{if } x \in V_2. \end{cases}$$

Then $\frac{\sum_{x \sim y} |f(x) - f(y)|}{\sum_x |f(x)|} = \frac{2}{n} \cdot \mathcal{F}(G_\sigma)$. ■

Therefore, in order to compute the frustration index of a signed graph, we need to solve the minimization problem of Theorem 3.7. To that end, we follow the approach described in [26], where the anti-symmetry constraint is accounted by solving the problem on one set of any admissible partition.

Theorem 3.8 *([26]). The global solution to the optimization problem $\min_{x \in D} \frac{f(x)}{g(x)}$, where $g(x) > 0, \forall x \in D$, is equivalent to the root of $h(\lambda) = \min_{x \in D}(f(x) - \lambda \cdot g(x))$, provided that $f(x) - \lambda g(x)$ is bounded below.*

Therefore the following alternating algorithm can be derived: Let $x_0 \in D$. Alternate the two following steps until convergence:

1. Compute $\lambda_k \frac{f(x_k)}{g(x_k)}$
2. Solve $x_{k+1} = \operatorname{argmin}_{x \in D}(f(x) - \lambda_k g(x))$.

It is shown in [26] that the algorithm decreases the objective function at each step and therefore converge to a (local) minima. As a starting vector, the first eigenvector of $L(G_\sigma)$ is used in practice.

We also notice that once a solution f of the above optimization problem is found, we can, in addition to the frustration index, find a set of edges solving for it by considering again the set of edges induced by $|\partial V_+(f)|$, where $V_+(f) = \{x \in \hat{V} | f(x) > 0\}$.

4. Real-life examples

The first example is a well-known social network describing relationships between cultures of the Central Highlands, in New Guinea [21]. The frustration index and the various bounds are shown in Table 1; while a solution to the frustration is given in Fig. 1. Interestingly, $\mathcal{F}_{l^2}(G_\sigma)$ is strictly greater than $\mathcal{F}(G_\sigma)$; showing that the frustration cannot be systematically computed using the first eigenvector of the signed Laplacian.

Our motivation for studying the frustration index is mainly linked to the suite of causal biological networks from [4]. Those 61 networks describe key processes such as cell fate, inflammation or cell stress. Similar networks have been used in [19,25], where a network scoring that requires network to be balanced is discussed. The frustrations and the vari-

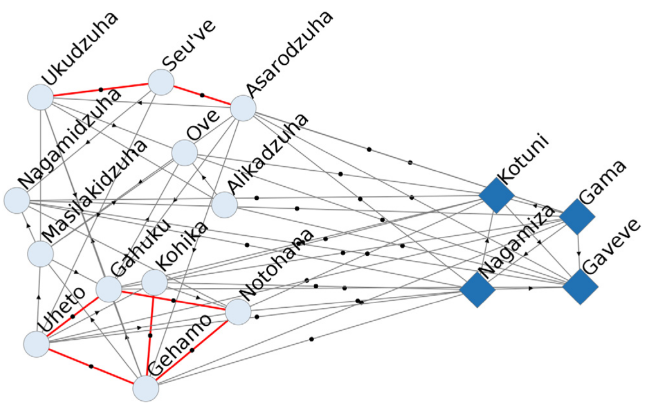

Fig. 1. One solution to the frustration problem for the cultures of the Central Highlands social network. Edges which removal leads to a balanced network are shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Frustration index, and its various bounds for the Central Highlands network. $f \in l^2(\hat{V})$ denotes the first eigenvector of the Laplacian for G_σ.

Index	Value
$\mathcal{F}(G_\sigma)$	7.00
$\frac{n}{4} \cdot \lambda_1(G_\sigma)$	4.16
$\mathcal{F}_{l^2}(G_\sigma)$	8.00
$\frac{n}{\sqrt{2}} \cdot \sqrt{\lambda_1(G_\sigma)(2\Delta - \lambda_1(G_\sigma))}$	50.25
$\frac{m-\sqrt{m}}{\sqrt{2}}$	35.63
$\min(\lvert E_+\rvert, \lvert E_-\rvert)$	29.00
$\lvert E\rvert$	58.00
$\lvert V\rvert$	16

ous bounds for this suite of networks are shown in Table 2. About one third of the networks are found to be balanced (19 out of 61).

The cell cycle and oxidative stress networks have the highest frustration. In [24], manual attempts to extract a maximal balanced sub-networks for those two networks in particular failed (while using closely related networks). Also none of the frustration was solved by removing $\min(|E_+|, |E_-|)$ edges.

In those examples, the upper spectral bounds are not tight when the graphs are unbalanced.

The closer upper bound is clearly obtained by using $\mathcal{F}_{l^2}(G_\sigma)$. We also observe that $\mathcal{F}(G_\sigma)$ is found by computing $\mathcal{F}_{l^2}(G_\sigma)$ for 28 out of the 42 unbalanced networks. As a consequence, as $\mathcal{F}_{l^2}(G_\sigma)$ is faster to compute, this bound is foreseen to be a useful approximation for very large graphs (hundreds of thousands of nodes).

5. Conclusion

Solving the frustration problem is combinatorial by nature and requires thereof efficient approaches. In this article, we established inequalities linking the frustration index and the first eigenvalue of the signed combinatorial Laplacian. From the main proof, we formulated the problem of finding a solution to the frustration as a l^1-norm optimization problem that can be efficiently implemented. This approach was illustrated on 61 biological networks.

From the proposed approach, a solution to remove the frustration in a graph is found. An interesting open problem is to enumerate several (at least many) solutions to the frustration in a given graph and, to our knowledge, remains unsolved.

Acknowledgments

The author would like to thank Pierre-Nicolas Jolissaint for his careful review and suggestions. Finally, I would like to thank the reviewers for numerous useful suggestions.

Table 2
Bounds and frustration index for the CausalBioNet network suite. Only the results for the 42 unbalanced networks are shown.

	$\mathcal{F}(G_\sigma)$	$\frac{n}{4}\cdot\lambda_1(G_\sigma)$	$\mathcal{F}_{l^2}(G_\sigma)$	$\frac{n}{\sqrt{2}}\cdot\sqrt{\lambda_1(G_\sigma)(2\Delta-\lambda_1(G_\sigma))}$	$\frac{m-\sqrt{m}}{\sqrt{2}}$	$\min(\lvert E_+\rvert, \lvert E_-\rvert)$	$\lvert E\rvert$	$\lvert V\rvert$
Cell Interaction	1	0.24	2	23.05	64.31	6	101	79
CV-IPN-Smooth muscle cell activation	1	0.41	1	49.97	117.11	13	179	137
Drug Metabolism Response	1	0.42	2	45.92	58.27	6	92	66
Mapk	1	0.07	2	8.46	32.99	12	54	44
Megakaryocyte Differentiation	1	0.39	1	47.68	102.83	14	158	122
Necroptosis	1	0.25	1	34.28	56.26	9	89	68
NK Signaling	1	0.16	1	17.41	49.58	2	79	66
Nuclear Receptors	1	0.14	1	13.06	25.77	8	43	37
Osmotic Stress	1	0.13	1	20.98	55.59	6	88	75
Cell Migration and Adhesion in Wound Healing	2	0.40	2	102.09	210.19	19	315	198
CV-IPN-Endothelial cell–monocyte interaction	2	0.48	2	54.94	71.04	5	111	87
Epithelial Innate Immune Activation	2	0.52	2	61.94	115.07	13	176	132
Epithelial Mucus Hypersecretion	2	0.59	2	94.18	136.20	23	207	134
Mast cell activation	2	0.57	2	38.80	40.92	4	66	55
Mechanisms of Cellular Senescence	2	0.44	2	75.10	120.52	19	184	139
Autophagy	3	0.79	3	109.07	130.74	31	199	135
ECM Degradation	3	0.97	3	60.21	80.48	24	125	67
Fibrosis	3	1.41	3	218.77	235.64	44	352	181
Wound Healing	3	1.17	3	170.96	212.25	19	318	196
Angiogenesis	4	1.46	4	174.56	216.37	44	324	238
Hypoxic Stress	4	1.43	4	133.83	124.60	36	190	125
mTor	4	1.28	4	39.59	36.29	7	59	44
Wnt	4	0.76	7	41.59	56.93	20	90	57
CV-IPN-Endothelial cell activation	5	1.73	5	205.78	272.15	36	405	245
Th1_Th2 Signaling	5	1.07	5	100.48	106.91	16	164	118
Xenobiotic Metabolism Response	5	1.65	5	99.04	79.81	9	124	71
CV-IPN-Plaque destabilization	6	2.48	6	241.60	331.52	72	491	245
Apoptosis	7	1.94	8	296.70	283.18	78	421	270
CV-IPN-Platelet activation	7	1.92	7	132.19	117.11	11	179	134
Endoplasmic Reticulum Stress	7	1.64	9	104.93	91.98	30	142	99
Endothelial Innate Immune Activation	7	1.86	8	127.64	127.33	23	194	137
Growth Factor	7	1.53	9	209.93	298.37	26	443	277
Microvascular endothelium activation	7	1.76	7	82.94	106.23	23	163	109
Clock	8	1.93	23	103.19	94.02	33	145	77
Macrophage Signaling	8	2.72	9	259.48	192.33	37	289	221
NFE2L2 Signaling	8	1.60	13	119.36	93.34	23	144	97
CV-IPN-Foam cell formation	9	1.58	9	134.74	188.90	33	284	192
Endothelial Shear Stress	9	2.21	9	147.97	90.63	23	140	92
Neutrophil Signaling	9	2.85	11	310.62	249.41	30	372	229
Response to DNA Damage	10	2.52	11	304.49	242.52	50	362	249
Cell Cycle	20	5.16	20	281.16	201.26	71	302	183
Oxidative Stress	21	5.20	28	579.77	341.89	66	506	337

References

[1] A.S. Bandeira, A. Singer, D.A. Spielman, A cheeger inequality for the graph connection laplacian, SIAM J. Matrix Anal. Appl. 34 (4) (2013) 1611–1630.
[2] F. Belardo, Balancedness and the least eigenvalue of laplacian of signed graphs, Linear Algebra Appl. 446 (2014) 133–147.
[3] Y. Bilu, N. Linial, Lifts, discrepancy and nearly optimal spectral gap, Combinatorica 26 (2006) 495–519.
[4] S. Boué, M. Talikka, J.W. Westra, W. Hayes, A. Di Fabio, J. Park, W.K. Schlage, A. Sewer, B. Fields, S. Ansari, F. Martin, E. Veljkovic, R. Kenney, M.C. Peitsch, J. Hoeng, Causal biological network database: A comprehensive platform of causal biological network models focused on the pulmonary and vascular systems, Database 2015 (2015) bav030.
[5] F.R. Chung, Spectral graph theory (cbms regional conference series in mathematics, no. 92), 1996.
[6] B. DasGupta, G. Enciso, E. Sontag, Y. Zhang, Algorithmic and complexity results for decompositions of biological networks into monotone subsystems, Biosystems 90 (1) (2007) 161–178.
[7] E. Estrada, M. Benzi, Are social networks really balanced?, 2014. J arXiv preprint arXiv:1406.2132.
[8] R. Figueiredo, Y. Frota, An improved branch-and-cut code for the maximum balanced subgraph of a signed graph, 2013.
[9] A. Gournay, An isoperimetric constant for signed graphs, Expo. Math. 34 (3) (2016) 339–351.
[10] F. Harary, On the notion of balance of a signed graph, Michigan Math. J. 2 (2) (1953) 143–146.
[11] F. Harary, J.A. Kabell, A simple algorithm to detect balance in signed graphs, Math. Social Sci. 1 (1) (1980) 131–136.
[12] Y.P. Hou, Bounds for the least laplacian eigenvalue of a signed graph, Acta Math. Sinica 21 (4) (2005) 955–960.
[13] Y.P. Hou, J. Li, Y. Pan, C. Dewey, On the laplacian eigenvalues of signed graphs, Linear Multilinear Algebra 51 (1) (2003).
[14] N.R. Huffner Falk, N. Betzler, Optimal edge deletions for signed graph balancing, in: Proc. 6th WEA, in: LNCS, vol. 4525, 2007, pp. 297–310.
[15] G. Iacono, C. Altafini, Monotonicity, frustration, and ordered response: an analysis of the energy landscape of perturbed large-scale biological networks, BMC Syst. Biol. 4 (2010) 83.
[16] G. Iacono, F. Ramezani, N. Soranzo, C. Altafini, Determining the distance to monotonicity of a biological network: a graph-theoretical approach, IET Syst. Biol. 4 (3) (2010) 223–235.
[17] H.-H. Li, C. Dewey, Note on the normalized laplacian eigenvalues of signed graphs, Aust. J. Combin. 44 (2009).

[18] A. Marcus, D.A. Spielman, N. Srivastava, Interlacing families ii: Mixed characteristic polynomials and the kadison-singer problem, 2013. J arXiv preprint arXiv:1306.3969.
[19] F. Martin, T. Thomson, A. Sewer, D. Drubin, C. Mathis, D. Weisensee, D. Pratt, J. Hoeng, M. Peitsch, Assessment of network perturbation amplitudes by applying high-throughput data to causal biological networks, BMC Syst. Biol. 6 (2012) 54.
[20] B. Mohar, Isoperimetric inequalities, growth, and the spectrum of graphs, Linear Algebra Appl. 103 (1988) 119–131.
[21] K.E. Read, Cultures of the central highlands, new guinea, Southwes. J. Anthropol. (1954) 1–43.
[22] E.D. Sontag, Monotone and near-monotone biochemical networks, Syst. Synth. Biol. 1 (2) (2007) 59–87.
[23] N. Soranzo, F. Ramezani, G. Iacono, C. Altafini, Decompositions of large-scale biological systems based on dynamical properties, Bioinformatics 28 (1) (2012) 76–83.
[24] T. Thomson, A. Sewer, F. Martin, V. Belcastro, B. Frushour, S. Gebel, J. Park, W. Schlage, M. Talikka, D. Vasilyev, J. Westra, J. Hoeng, M. Peitsch, Quantitative assessment of biological impact using transcriptomic data and mechanistic network models, Toxicol. Appl. Pharmacol. 272 (3) (2013) 863–878.
[25] D. Vasilyev, T. Thomson, B. Frushour, F. Martin, A. Sewer, An algorithm for score aggregation over causal biological networks based on random walk sampling, BMC Res. Notes 7 (2014) 516.
[26] H. Wang, F. Nie, H. Huang, Robust distance metric learning via simultaneous l1-norm minimization and maximization, in: ICML, 2014, pp. 1836–1844.
[27] P.-S. Wei, B.Y. Wu, Balancing a complete signed graph by changing minimum number of edge signs, in: 29th Workshop on Combinatorial Mathematics and Computation Theory, 2012.
[28] T. Zaslavsky, Signed graphs, Discrete Appl. Math. 4 (1) (1982) 47–74.
[29] T. Zaslavsky, Balanced decompositions of a signed graph, J. Combin. Theory Ser. B 43 (1) (1987) 1–13.