An exact method for computing the frustration index in signed networks using binary programming

Samin Aref, Andrew J. Mason*, and Mark C. Wilson
Department of Computer Science *Department of Engineering Science
University of Auckland, Auckland, Private Bag 92019, New Zealand
Email: sare618@aucklanduni.ac.nz

November 28, 2016

Abstract

Computing the frustration index of a signed graph is a key to solving problems in different fields of research including social networks, physics, material science, and biology. In social networks the frustration index determines network distance from a state of structural balance. Although the definition of frustration index goes back to 1960, an exact algorithmic computation method has not yet been proposed. The main reason seems to be the complexity of computing the frustration index which is closely related to well-known NP-hard problems such as MAXCUT.

New quadratic and linear binary programming models are developed to compute the frustration index exactly. Using the Gurobi solver, we evaluate the frustration index on real-world and synthetic datasets. The synthetic data involves Erdős-Rényi networks, Barabási-Albert networks, and specially structured random graphs. We also use well-known datasets from the sociology literature, such as signed networks inferred from students' choice and rejection as well as datasets from the biology literature including gene regulatory networks. We also provide some results on the frustration index of a political network of countries over time.

The results show that exact values of the frustration index can be efficiently computed using our suggested optimisation models. We find that most real-world social networks and some biological networks exhibit a relatively low level of frustration which indicates that they are close to balanced.

Keywords: Integer programming, Optimisation, Frustration index, Branch and bound, Signed graphs, Balance theory

Mathematics Subject Classification (MSC 2010):
90C10 90C20 90C35 90C57 90C90 05C15 11E16 65K05

1 Introduction

Local ties between entities lead to global structures in networks. Ties can be formed as a result of interactions and individual preferences of the entities in the network. The dual nature of interactions in various contexts means the ties may form in two opposite types, namely positive ties and negative ties. In a social context, this is interpreted as friendship vs. enmity or trust vs. distrust between people. The term *signed network* embodies a multitude of concepts involving relationships characterisable by ties with plus and minus signs. *Signed graphs* are used to model such networks where edges have positive and negative signs. *Structural balance* in signed graphs is a macro-scale structural property that has recently become a focus in network science analytical models.

Originated from social psychology, structural balance theory was the first attempt to understand the sources of tensions and conflicts in groups of people with signed ties [1]. Cartwright and Harary identified cycles of the graph (closed-walks with distinct nodes) as the origins of tension, in particular cycles containing an odd number of negative edges [2]. Signed graphs in which no such cycles are present hold the property of structural balance. The vertex set of *balanced* signed networks can be partitioned into $k \leq 2$ subsets such that each negative edge joins vertices belonging to different subsets [2]. For graphs that are not totally balanced, a distance from total balance (a measure of partial balance) can be computed. Among various measures is the *frustration index* that indicates the minimum number of edges whose removal (or equivalently, negation) results in total balance [3].

2 Literature review

In the past few decades, different measures of balance [2, 4–7] are suggested and deployed to analyse partial balance in real-world signed networks resulting in conflicting observations. Measures based on cycles [2,4], triangles [5,6] and closed-walks [7] are not generally consistent and do not hold axiomatic properties [8]. Among all the measures, a normalised version of frustration index is shown to satisfy basic axioms (listed in 3.4) [8]. This measure provides a clear understanding of the transition to total balance in terms of the number of edges to be modified and reduce the tension as opposed to graph cycles that were first suggested as origins of tension in unbalanced networks.

The frustration index is a key to frequently stated problems in many different fields of research. In biological networks, optimal decomposition of network into monotone subsystems -which is essential for understanding Drosophila segment polarity- is made possible by calculating the signed graph frustration index [9]. In finance, risk in a portfolio of securities can be managed by analysing the frustration index of signed graphs [10]. In Knot theory, the Writhe of a link diagram is a measure directly related to the frustration index [11]. In material science, the frustration index provides the minimum energy state of magnetic materials [12]. In international relations, signed clustering of countries in a region can be investigated using the frustration index [13]. In nano-materials, bipartite edge frustration has applications on the stability of fullerene, a carbon allotrope [14]. Finally in electronics, frustration index helps finding the minimal set of phase conflicts in integrated circuit design [15].

From a computational viewpoint, detecting if a graph is balanced is not a difficult problem and can be solved in polynomial time [16]. However, calculating the frustration index is an NP-hard problem equivalent to the ground state calculation of the Sherrington-Kirkpatrick spin glass model [17, 18]. Computation of frustration index can be also reduced from the graph maximum cut (MAXCUT) problem, in a special case of all negative edges, which is known to be NP-hard [16]. MAXCUT can be solved in polynomial time for planar graphs [19] and efficient approximations have long existed for MAXCUT in general graphs [20]. While the frustration index can be computed in polynomial time for planar graphs [21], the frustration index in general graphs, which is as difficult as the weighted version of MAXCUT with negative weights, lacks extensive and systematic investigation.

Despite the lack of investigation of the frustration index, a closely-related and more general problem in signed graphs has been investigated comprehensively. According to Davis's definition of *generalized balance*, a signed network is *weakly balanced* (k-balanced) iff its vertex set can be partitioned into k subsets such that each negative edge joins vertices belonging to different subsets [22]. The problem of finding the minimum number of frustrated edges for general k (an arbitrary number of subsets) is referred to as the *Correlation Clustering* problem. For every fixed k, there is a polynomial time approximation scheme for the correlation clustering problem [23]. For arbitrary k, exact [24,25] and heuristic methods [26,27] are developed based on a mixed integer programming model [28]. Denoting the order of a graph by n, exact algorithms fail for $n > 21$ [24] and $n > 40$ [25], while greedy algorithms [26] and local search heuristics [27] are capable of providing good solutions for $n \leq 10^3$ and $n \leq 10^4$.

Doreian and Mrvar have recently analysed signed international relations [29] using the frustration index and correlation clustering [13] arguing that the frustration index is computable in polynomial time. Their claims must be evaluated while maintaining a certain level of healthy skepticism.

For k fixed to 2, efficient data reduction schemes [30] and ground state search heuristics [9] are suggested that provide bounds for the frustration index. Iacono et al. showed that the frustration index equals the minimum number of fundamental negative cycles over all spanning trees of the graph [9]. Originally discussed in the biology context, the terminology used in [9] is different where *monotonocity* and *inconsistency* are the equivalences for balance and frustration. Facchetti, Iacono, and Altafini suggested a non-linear energy function minimization model for finding the frustration index [31]. Their model was solved using various techniques [9, 32–34]. Using the ground state search heuristic algorithms [9], the frustration index is estimated in biological networks up to $n \leq 1.5 \times 10^3$ [9] and social networks up to $n \leq 10^5$ [31, 35].

After extending the non-linear energy minimization model to weak structural balance, Ma et al. provided good solutions for the correlation clustering problem in networks up to $n \leq 10^5$ using various heuristics [33, 34]. Esmailian et al. have also extended the work of Facchetti, Iacono, and Altafini, focusing on the role of negative ties in signed graph clutering [32, 36]. Their suggested heuristic is reported to solve networks with up to $n \leq 10^5$ within 99% of optimality. However, not only their main theorem (Theorem 1 in [32]) is incorrect, but Mendonça et al. has also cast doubt on their main conclusion regarding the role of negative ties in signed graphs [37]. The analysis of literature shows that there is a critical gap in the exact methods for computing the frustration index that can guarantee the solution quality.

Our contribution

The principal focus of this research study is to provide insight into computing the frustration index. Besides multiple applications in various fields of research, another motivation for computing this measure is to systematically investigate signed networks transition to balance using basic graph operations on frustrated edges. Thus we develop the first algorithmic method for exact computation of the frustration index.

Starting with a quadratic programming model based on signed graph switching equivalences, we suggest several optimisation models. The advantage of formulating the problem as an optimisation model is not only exploring the details involved in a fundamental NP-hard problem, but making use of powerful mathematical programming solvers like Gurobi to solve the NP-hard problem exactly and efficiently. This approach allows us to compute a measure overlooked for decades because of the inherent combinatorial complexity. We provide a numerical comparison of the frustration index on a variety of undirected signed networks, both randomly generated and inferred from well-known datasets.

This paper begins by laying out the theoretical dimensions of the research in Section 3. A quadratic programming model is developed in Section 4 and linear programming models are formulated in Section 5. The results on synthetic data are provided in Section 6 which also contains closed-form formula for the frustration index in specially structured graphs and discussions on the performance of the binary linear models. Numerical results on real social and biological networks are provided in Section 7. The frustration index of a temporal network is provided in Section 8. Section 9 presents the findings of the research and sums up the research highlights.

3 Preliminaries

3.1 Basic notation

We consider undirected signed networks $G = (V, E, \sigma)$. The set of nodes is denoted by V, with $|V| = n$. E is the set of edges that is partitioned into the set of positive edges E^+ and the set of negative edges E^- with $|E| = m$, $|E^-| = m^-$, and $|E^+| = m^+$ where $m = m^- + m^+$. σ is the sign function $\sigma : E \to \{-1, +1\}$. The adjacency matrix $\mathbf{A}$ is defined in (1).

$$\mathbf{A}_{ij} = \begin{cases} \sigma_{i,j} & \text{if } (i,j) \in E \\ 0 & \text{if } (i,j) \notin E \end{cases} \tag{1}$$

The number of positive (negative) edges connected to the node $i \in V$ represent positive (negative) degree of the node and is denoted by $d^+(i)$ ($d^-(i)$). The net degree of a node is defined by $d^+(i) - d^-(i)$.

A *walk* of length k in G is a sequence of nodes $v_0, v_1, ..., v_{k-1}, v_k$ such that for each $i = 1, 2, ..., k$ there is an edge from v_{i-1} to v_i. If $v_0 = v_k$, the sequence is a *closed walk* of length k. If the nodes in a closed walk are distinct except endpoints, it is a directed cycle (for simplicity *cycle*) of length k. The *sign* of a cycle is the product of the signs of its edges and cycles with negative signs are unbalanced. A balanced cycle is one with positive sign. A balanced graph is one with no negative cycles.

3.2 Node colouring and frustration count

Satisfied and *frustrated* edges are defined based on two-colourings of the nodes. Colouring the nodes with black and white, a frustrated (satisfied) edge (i, j) is either a positive (negative) edge with different colours on the endpoints i, j or a negative (positive) edge with the same colours on the endpoints i, j.

To be more specific, for any signed graph $G = (V, E, \sigma)$, we can partition V into two sets, denoted $X \subseteq V$ and $\bar{X} = V \backslash X$. We think of X as specifying a colouring of the nodes, where each node $i \in X$ is coloured black, and $i \in \bar{X}$ is coloured white. We let x_i denote the colour of node $i \in V$ under X, where $x_i = 1$ if $i \in X$ and $x_i = 0$ otherwise. We say that an edge (i, j) is *frustrated* under X if either edge (i, j) is a positive edge (i.e. $(i, j) \in E^+$) but nodes i and j have different colours ($x_i \neq x_j$), or edge (i, j) is a negative edge (i.e. $(i, j) \in E^-$) but nodes i and j share the same colour ($x_i = x_j$). We define the *frustration count* $f_G(X)$ as the number of frustrated edges under X:

$$f_G(X) = \sum_{(i,j) \in E} f_{i,j}(X)$$

where

$$f_{i,j}(X) = \begin{cases} 0, & \text{if } x_i = x_j \text{ and } (i,j) \in E^+ \\ 1, & \text{if } x_i = x_j \text{ and } (i,j) \in E^- \\ 0, & \text{if } x_i \neq x_j \text{ and } (i,j) \in E^- \\ 1, & \text{if } x_i \neq x_j \text{ and } (i,j) \in E^- \end{cases} \tag{2}$$

The frustration index $L(G)$ of a graph G can be found by finding a subset $X^* \subseteq V$ of G that minimises the frustration count $f_G(X)$, i.e. solving Eq. (3).

$$\min_{X \subseteq V} f_G(X) \tag{3}$$

3.3 Deletion-minimal set and switching function

For each signed graph, there is a collection of edges $E^* \subseteq E$, called *deletion-minimal*, whose deletion results in a balanced graph while there is no subset of this collection that yields balance if deleted. The frustration index $L(G)$ equals the number of members in a deletion-minimal set: $L(G) = |E^*|$.

We define the *switching function* g operating over a set of vertices, called the *switching set*, $X \subseteq V$ as follows in (4).

$$\sigma^g_{(i,j)} = \begin{cases} \sigma_{i,j} & \text{if } i,j \in X \text{ or } i,j \notin X \\ -\sigma_{i,j} & \text{if } i \in X \text{ and } j \notin X \text{ or } i \notin X \text{ and } j \in X \end{cases} \tag{4}$$

The graph resulted from applying switching function g to signed graph G is called G's *switching equivalence* and denoted by G^g. The switching equivalences of a graph have the same value of the frustration index $\forall g : L(G^g) = L(G)$ [16]. $\mathbf{A}^g$ represents the adjacency matrix of the graph switched by g. It is straightforward to prove that the frustration index is equal to the minimum number of negative edges over all switching functions [16]. An immediate result is that any balanced graph can switch to an equivalent graph where all the edges are positive [16]. Moreover, in the switched graph with the minimal negative edges, called *negative minimal graph* and denoted by G^{g^*}, all vertices have a non-negative net degree. In other words, if $m^-(G^g) = L(G)$ then every vertex $\forall i \in V$ under switching g satisfies $d^-(i^g) \leq d^+(i^g)$.

To see that Eq. (3) is equivalent to finding a negative minimal graph, we note that any $X \subseteq V$ defines a switching set that results in a switched graph $G^g = (V, E, \sigma^g)$ in which $\sigma^g_{i,j} = -1$ if and only if edge $(i,j) \in E$ in our original graph G is frustrated under X. Therefore, the count of negative edges in G^g equals the frustration count $f_G(X)$, and so $L(G) = f_G(X^*)$. Note that $f_G(X)$ gives an upper bound on $L(G)$ for any $X \subseteq V$.

3.4 Axioms for a measure of balance

The level of partial balance in signed graphs can be measured by several methods. The lack of standard in the suggested methods based on cycles [2, 4], triangles [5, 6], matrix eigenvalues [38], and closed-walks [7] led Aref and Wilson to suggest the following axioms for a measure of partial balance [8]:

A1 A measure of balance is bounded by 0 and 1.

A2 Only total balance is denoted by value 1.

A3 Overall balance of two graphs is bounded between their individual balance.

A4 Adding a balanced cycle increases balance.

A5 Switching nodes does not change balance.

A6 Removing a deletion-minimal edge does not decrease balance.

A7 Adding a frustrated edge does not increase balance.

Among all the measures, the normalised frustration index, $F(G) = 1 - 2L(G)/m$, satisfies the axioms [8] and therefore is suggested for measuring and comparing the level of partial balance in signed networks.

3.5 Upper bounds for the frustration index

Upper bounds can be readily provided for the frustration index such as $L(G) \leq m^-$ which states the obvious result that removing all negative edges gives a balanced graph. Tomescu [39] proves that frustration index is bounded by $\left\lfloor (n-1)^2/4 \right\rfloor$. Bounds for the largest frustration count for a given graph (over all colourings) are provided by [40]:

$$\frac{m}{2} - \sqrt{mn} \leq \max f_G(X) \leq \frac{m}{2}. \tag{5}$$

4 Quadratic programming models

In this section, we formulate two quadratic programming models in (6) and (9) to calculate the frustration index by minimising an objective function formed using variables defined over the nodes of the graph.

4.1 A quadratically constrained quadratic programming model

We start by developing a mathematical programming model in Eq. (6) to maximise Z_1 the sum of entries of $\mathbf{A}^g$ over different switching functions g. Bearing in mind that the frustration index is the number of negative edges in the negative minimal graph, $L(G) = m^-_{(G^{g^*})}$, the maximising Z_1 will calculate the frustration index. Decision variables, $y_i \in \{-1, 1\}$ define the black coloured nodes $X = \{i | y_i = 1\}$ (alternatively nodes in the switching set). The restriction for the variables is formulated by n quadratic constraints $y_i^2 = 1$. Note that the switching set $X = \{i | y_i = 1\}$ creates the negative minimal graph with the adjacency matrix entries given by $a_{ij} y_i y_j$. Maximising $\sum_{i \in V} \sum_{j \in V} a_{ij} y_i y_j$ is equivalent to minimising $m^-_{(G^{g^*})} = |(i,j) \in E : a_{ij} y_i y_j = -1|$. The model can be represented as Eq. (6) in the form of a continuous quadratically constrained quadratic programming (QCQP) model with n decision variables and n constraints.

$$
\begin{aligned}
\max_{y_1, y_2, \dots} \quad & Z_1 = \sum_{i \in V} \sum_{j \in V} a_{ij} y_i y_j \\
\text{s.t.} \quad & y_i^2 = 1 \quad \forall i \in V
\end{aligned} \tag{6}
$$

The optimal value of the objective function Z_1^* is equal to the sum of entries in the adjacency matrix of the negative minimal graph which can be represented by $Z_1^* = 2m^+_{(G^{g^*})} - 2m^-_{(G^{g^*})} = 2m - 4L(G)$. Therefore, the graph frustration index can be calculated by $L(G) = (2m - Z_1^*)/4$.

While the model expressed in (6) is quite similar to the non-linear energy function minimization model used in [32, 33, 35, 41] and the Sherrington-Kirkpatrick spin glass model [17], the feasible region in model (6) is neither convex nor a second order cone. Therefore, model (6) only serves as an easy-to-understand optimisation model clarifying the node colouring (alternatively selecting nodes to switch) and how it relates to the frustration index.

4.2 An unconstrained binary quadratic programming model

The optimisation model (6) can be converted into an unconstrained binary quadratic programming (UBQP) model (9) by changing the decision variables into binary variables $y_i = 2x_i - 1$ where $x_i \in \{0, 1\}$. Note that the binary variables, x_i, define the black coloured nodes $X = \{i | x_i = 1\}$ (alternatively nodes in the switching set). The optimal solution represents a subset $X^* \subseteq V$ of G that minimises the frustration count.

Furthermore, the terms in the objective function can be modified as shown in (7)–(8) in order to directly represent $L(G)$ in the objective function Z_2.

$$
\begin{aligned}
Z_1 &= \sum_{i\in V}\sum_{j\in V}(4a_{ij}x_ix_j - 2x_ia_{ij} - 2x_ja_{ij} + a_{ij}) \\
&= \sum_{i\in V}\sum_{j\in V}(4a_{ij}x_ix_j - 4x_ia_{ij}) + (2m - 4m^-_{(G)})
\end{aligned} \tag{7}
$$

$$
Z_2 = (2m - Z_1)/4 \tag{8}
$$

Note that the binary quadratic model in Eq. (9) has n decision variables and no constraints.

$$
\begin{aligned}
\min_{x_1,x_2,\dots} Z_2 = \sum_{i\in V}\sum_{j\in V} a_{ij}x_i(1-x_j) + m^-_{(G)} \\
\text{s.t.} \quad x_i \in \{0,1\}
\end{aligned} \tag{9}
$$

The optimal value of the objective function in Eq. (9) is denoted by Z_2^* which represents the frustration index directly as shown in (10).

$$
Z_2^* = (2m - Z_1^*)/4 = (2m - (2m - 4L(G)))/4 = L(G) \tag{10}
$$

Note that the term $a_{ij}x_i(1-x_j)$ in (9) takes value 1 for a positive frustrated edge and value -1 for a negative satisfied edge. Therefore, the objective function in Eq. (9) can be interpreted as initially starting by $m^-_{(G)}$ and then adding 1 for each positive frustrated edge and -1 for each negative satisfied edge which results in the total number of frustrated edges.

Having discussed the continuous quadratic (6) and binary quadratic (9) programming models, the next section of this paper addresses linear programming models as well as the structural properties of the problem.

5 Linear programming models

In this section, we formulate three linear programming models in (11), (17), and (19).

5.1 The AND model

The linearised version of (9) is the AND model. As formulated in Eq. (11), the non-linear term x_ix_j in the objective function of Eq. (9) is replaced by additional binary variables for each edge x_{ij} that take value 1 whenever $x_i = x_j = 1$ and 0 otherwise. Note that the binary variables taking value 1, x_{ij}, represent edges for which the both endpoints are in the switching set. The dependencies between the x_{ij} and x_i, x_j values are taken into account by considering a constraint for each new variable. Therefore, the binary linear formulation of the model has $n + m$ variables and m constraints as it follows in (11).

$$\min_{x_1,x_2,\ldots,x_{12},x_{13},\ldots} Z_2 = \sum_{i\in V}\sum_{j\in V}(a_{ij}x_i - a_{ij}x_{ij}) + m^-_{(G)}$$

$$\begin{aligned} \text{s.t.}\quad & x_{ij} \leq (x_i + x_j)/2 \quad \forall (i,j) \in E^+ \\ & x_{ij} \geq x_i + x_j - 1 \quad \forall (i,j) \in E^- \\ & x_i, x_{ij} \in \{0,1\} \end{aligned} \tag{11}$$

Having discussed the first linear programming formulation of the problem, we move on to the structural properties of the problem to restrict the feasible space.

The structural properties of the model allow us to restrict the model by adding additional valid inequalities. Properties of the the optimal solution can be used to determine additional constraints. Properties observed in negative minimal graphs include the nonnegativity of a node's net degree and negation states of the edges making a cycle.

An obvious structural property of the nodes in the negative minimal graph, G^{g^*}, is that their net degrees are always non-negative, i.e. $d^+(i^{g^*}) - d^-(i^{g^*}) \geq 0 \forall i \in V$. Equivalently, a node i should be given a colour that minimises the number of frustrated edges connected to it. This can be proved by contradiction using the definition of the switching function (4).

Proof. Assume a node in the negative minimal graph has a negative degree. It follows that the negative edges connected to the node outnumber the positive edges. Therefore, switching the node decreases the total number of negative edges in the negative minimal graph which is a contradiction. □

This structural property can be formulated as constraints in the problem. *Net-degree constraint* can be added to the model for each node restricting all variables associated with the connected edges. They are formulated using quadratic terms of x_i variables. As x_i represents the colour of a node, $(1-2x_i)(1-2x_j)$ takes value -1 if and only if the two endpoints of edge (i,j) have different colours. It means the edge should be negated in the process of transforming to the negative minimal graph. The quadratic formulation of our net-degree constraints is provided in (12). The linearised version using x_i and x_{ij} variables is provided in (13).

$$\sum_{j\in V} a_{ij}(1-2x_i)(1-2x_j) \geq 0 \quad \forall i \in V \tag{12}$$

$$\sum_{j\in V} a_{ij}(1-2x_i-2x_j+4x_{ij}) \geq 0 \quad \forall i \in V \tag{13}$$

Another structural property observed is related to the edges making a cycle. According to the definition of switching function (4), switching one node negates all edges connected to that node. As in a cycle, there are two edges connected to each node, the negation states of edges making a cycle are not independent. To be more specific, the number of negated edges in each cycle of the graph must be even in G^{g^*}.

As listing all cycles of a graph is computationally intensive, this structural property can be applied to cycles of a limited length. For instance, we may apply this structural property to the edge variables making triangles in the graph. In this case, for each triangle, half of the combination of variable values is cut from the feasible space. This structural property can be formulated as valid inequalities in Eq.(14) in which $T = \{(i, j, k) \in V^3 | (i, j), (i, k), (j, k) \in E\}$ denotes the set of nodes making a triangle. Note that, $(x_i + x_j - 2x_{ij})$ represents the negation state of the edges (i, j). The expression in Eq.(14) denotes the sum of negation states for the three edges ij, ik, jk making a triangle.

$$\begin{aligned} x_i + x_j - 2x_{ij} + x_i + x_k - 2x_{ik} + x_j + x_k - 2x_{jk} \\ = 0 \text{ or } 2 \quad \forall (i, j, k) \in T \end{aligned} \tag{14}$$

Eq.(14) can be linearised to Eq.(15) as follows. *Triangle constraints* can be applied to the model as four constraints per triangle restricting three edge variables and three node variables per triangle.

$$\begin{aligned} x_i + x_{jk} &\geq x_{ij} + x_{ik} \quad \forall (i, j, k) \in T \\ x_j + x_{ik} &\geq x_{ij} + x_{jk} \quad \forall (i, j, k) \in T \\ x_k + x_{ij} &\geq x_{ik} + x_{jk} \quad \forall (i, j, k) \in T \\ 1 + x_{ij} + x_{ik} + x_{jk} &\geq x_i + x_j + x_k \quad \forall (i, j, k) \in T \end{aligned} \tag{15}$$

More restrictions can be imposed using the minimum of the two frustration upper bounds, $Z_2 \leq \min\{m/2, m^-\}$ as discussed in 3.5. They are implemented as constraints in Eq. (16).

$$Z_2 \leq \min\{m/2, m^-\} \tag{16}$$

The complete formulation of the AND model with further restrictions on the feasible space includes the objective function and core constraints in Eq. (11) and valid inequalities in Eq. (13), Eq. (15), and Eq. (16). The model has $n + m$ binary variables, m core constraints, and $n + 4|T| + 1$ additional constraints. The next subsection discusses an alternative binary linear model for calculating the frustration index.

5.2 The XOR model

The problem of finding the frustration index can be directly formulated as a binary linear model. The XOR model is designed to directly count the number of edges whose negation leads to a negative minimal graph. We define binary variable u_{ij} $\forall (i,j) \in E$ denoting negation state of each edge. u_{ij} takes value 1 if the edge is negated in the negative minimal graph, otherwise it takes value 0. We use x_i $\forall i \in V$ denoting the colouring of a node similar to the previous models.

Therefore, the frustration count under a node colouring $(x_1, x_2, \ldots, x_n)$ is given by:

$$\sum_{(i,j)\in E^+} u_{ij} + \sum_{(i,j)\in E^-} (1 - u_{ij}) = \sum_{(i,j)\in E} a_{ij}u_{ij} + m^-_{(G)}$$

The respective minimisation model is as follows in Eq. (17).

$$\begin{aligned} \min_{x_1,x_2,\ldots,u_{12},u_{13},\ldots} Z_3 = \sum_{(i,j)\in E} a_{ij}u_{ij} + m^-_{(G)} \\ \text{s.t.} \quad \text{core constraints} \\ u_{ij} \leq 2 - x_i - x_j \forall (i,j) \in E^- \\ u_{ij} \leq x_i + x_j \quad \forall (i,j) \in E^- \\ u_{ij} \geq x_i - x_j \quad \forall (i,j) \in E^+ \\ u_{ij} \geq x_j - x_i \quad \forall (i,j) \in E^+ \\ x_i, u_{ij} \in \{0,1\} \end{aligned} \tag{17}$$

The objective function in (17) can be interpreted as starting by the initial number of negative edges then adding 1 for each negated positive edges and adding -1 for each negated negative edges.

The counterparts of the valid inequalities (13), (15), and (16) for the XOR model is formulated in Eq. (18) using the u_{ij} variables.

The complete formulation of the XOR model includes the objective function and core constraints in Eq. (17) and valid inequalities in Eq. (18) which has $n+m$ binary variables, $2m$ core constraints, and $n+4|T|+1$ additional constraints.

$$\begin{aligned} \text{valid inequalities} \\ \sum_{j\in V} a_{ij}(1 - 2u_{ij}) \geq 0 \quad \forall i \in V \\ u_{ij} + u_{ik} \geq u_{jk} \quad \forall (i,j,k) \in T \\ u_{ij} + u_{jk} \geq u_{ik} \quad \forall (i,j,k) \in T \\ u_{ik} + u_{jk} \geq u_{ij} \quad \forall (i,j,k) \in T \\ u_{ij} + u_{ik} + u_{jk} \leq 2 \quad \forall (i,j,k) \in T \\ Z_3 \leq \min\{m/2, m^-\} \end{aligned} \tag{18}$$

A third linear formulation of the problem is provided in the next subsection.

5.3 The ABS model

In this section, we propose an alternative binary linear model in which edge variables represent the frustration state of the edges.

We start by observing that for a node colouring $(x_1, x_2, \dots, x_n)$, $|x_i - x_j| = 1$ for a positive frustrated edge and $|x_i - x_j| = 0$ for a positive satisfied edge $(i, j) \in E^+$. Similarly, $1 - |x_i - x_j|$ gives the frustration state of a negative edge $(i, j) \in E^-$. Introducing two binary variables e_{ij}, f_{ij} for each edge that collectively represent the frustration state of the edge (i, j) allows us to formulate the linear model in (19).

The objective function, being the total number of frustrated edges, sums the above mentioned absolute value terms to compute the frustration count in (19). Linear constraints expressed in (19) account for determining the frustration state of an edge based on the node variables x_i, x_j that represent the two endpoint colourings of the edge (i, j).

$$
\begin{aligned}
\min_{x_1,x_2,\dots,e_{12},e_{13},\dots,f_{12},f_{13},\dots} \quad & Z_4 = \sum_{(i,j)\in E} e_{ij} + f_{ij} \\
\text{s.t.} \quad x_i + x_j - 1 = \quad & e_{ij} - f_{ij} \quad \forall (i,j) \in E^- \\
x_i - x_j = \quad & e_{ij} - f_{ij} \quad \forall (i,j) \in E^+ \\
& x_i, e_{ij}, f_{ij} \in \{0, 1\}
\end{aligned}
\tag{19}
$$

In order to speed up the model in (19), we consider adding a constraint to increase the root node objective function. The linear programming relaxation of the model in (19), sets all node variables to 0.5 resulting in all edge variables staying at 0 which minimises the objective function and gives a root node objective of value 0. However, by fixing one node variable, we can increase the root node objective by breaking the symmetry that exists and allows switching the colour on each node to give an equivalent solution. We conjecture the best node variable is the one associated with the highest unsigned node degree. This constraint is formulated in (20) which our experiments show speeds up the branch and bound algorithm by increasing the lower bound.

$$u_k = 1 \quad d_k = \max_{i \in V} d_i \tag{20}$$

Based on the same concept, we may configure the branch and bound algorithm so that it branches first on the node with the highest unsigned degree.

As another effort to make the algorithm more effective, we consider adding valid inequalities to the model in (19). According to [16], every unbalanced cycle of the graph contains an odd number of frustrated edges. This means that any colouring of the nodes in an unbalanced triangle must produce at least one frustrated edge. Recalling that $e_{ij} + f_{ij}$ is 1 if edge (i, j) is frustrated (and 0 otherwise), then for any node triple (i, j, k) defining an unbalanced triangle in G, we have the valid inequality (21).

$$e_{ij} + f_{ij} + e_{ik} + f_{ik} + e_{jk} + f_{jk} \geq 1 \quad \forall (i, j, k) \in T^- \tag{21}$$

In (21), $T^- = \{(i, j, k) \in V^3 | \sigma_{(i,j)}\sigma_{(i,k)}\sigma_{(j,k)} = -1\}$ denotes the set of nodes defining an unbalanced triangle. The expression in Eq. (21) denotes the sum of frustration states for the three edges $(i, j), (i, k), (j, k)$ making an unbalanced triangle.

The complete formulation of the alternative binary linear programming model includes the objective function and core constraints in Eq. (19) – (20) and valid inequalities in Eq. (21). The model has $n + 2m$ binary variables, m core constraints, and $|T^-|$ valid inequalities.

5.4 Comparison of the five models

In this subsection we compare the five models based on the number and type of variables and constraints. Table 1 summarises the five models developed in Sections 4 and 5.

Table 1: Comparison of the variables and constraints in the five models

	QCQP (6)	UBQP (9)	AND (11)	XOR (17)	ABS (19)
Variables	n	n	$n + m$	$n + m$	$n + 2m$
Constraints	n	0	m	$2m$	m
Variable type	continuous	binary	binary	binary	binary
Constraint type	quadratic	-	linear	linear	linear
Objective	quadratic	quadratic	linear	linear	linear

As Table 1 shows, the number and types of variables and constraints differ in the five models. While one can suggest ways to transform one model to the other which indicates their mathematical equivalence, they perform differently in terms of solve time, root node objective and the number of branch and bound (B&B) nodes required to solve a given problem.

Valid inequalities are utilised as additional non-core constraints that are kept aside from the core constraints of the model. Upon violation by a solution, valid inequalities are efficiently pulled in to the model. Pulled in valid inequalities cut a part of the feasible space and restrict the model. Additional restrictions imposed on the model would speed up the solver algorithm if they are valid and useful [42].

The names used for the three linear models reflect their model formulation concepts. The AND model features x_{ij} variable denoting the AND function of the two binary variable x_i, x_j. Similarly, u_{ij} variable in the XOR model and e_{ij}, f_{ij} variables in the ABS model represent the XOR function and the absolute difference of variables x_i, x_j.

Having discussed various programming models and speed-up techniques for computing the frustration index, the next section presents the numerical results on synthetic data using the proposed models.

6 Numerical results in random graphs

In this section, the frustration index of various random networks is computed by solving the optimisation models using Gurobi version 6.5.1 on a desktop computer with an Intel Corei5 4670 @ 3.40 GHz and 8.00 GB of RAM running 64-bit Microsoft Windows 7. The models were created using the Gurobi Python environment.

6.1 Small random signed graphs

Both Erdős-Rényi and Barabási-Albert random graphs are used as synthetic data for calculation of the frustration index. In this analysis, the same randomly generated graphs with different numbers of negative edges assigned by a uniform random distribution are used as test cases over 50 runs per experiment setting. Figure 1 demonstrates the average and standard deviation of frustration index in these random signed networks with $n = 15, m = 50$. It is worth mentioning that similar results are observed in other types of random graphs including small world, scale-free, and random regular graphs.

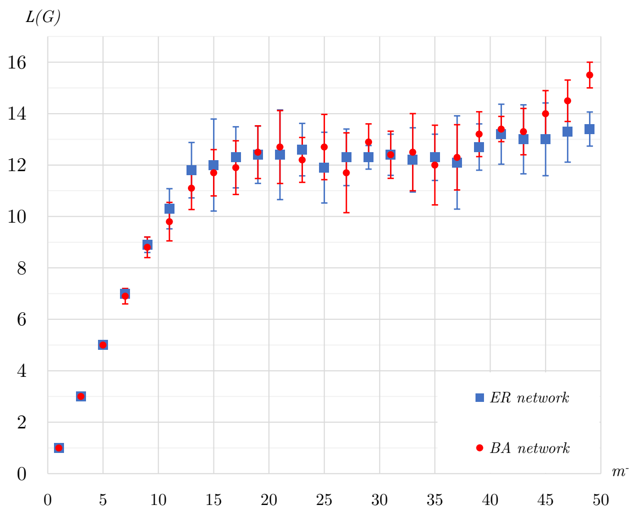


Figure 1: (colour online) The frustration index in Erdős-Rényi random graphs with 15 nodes and 50 edges ($n = 15, p = 0.45$) and Barabási-Albert random graphs with 15 nodes and 50 edges and various number of negative edges

Figure 1 shows similar increases in the frustration index in the two graph classes as m^- increases. It can be observed that the maximum frustration index is still smaller than $m/3$. This shows a gap between the values of the frustration index in random graphs and the theoretical upper bound of $m/2$. It is important to know whether this gap is proportional to graph size and density.

6.2 4-regular signed graphs of different orders

In order to investigate the gap discussed earlier, 4-regular random graphs with a constant fraction of randomly assigned negative edges are analysed averaging over 50 runs per experiment setting. The frustration index is computed for 4-regular random graphs with 25%, 50%, and 100% negative edges and compared with the upper bound $m/2$. Figure 2 demonstrates the average and standard deviation of frustration index where the degree of all nodes remains constant, but the the density decreases as the graph grows in size and order.

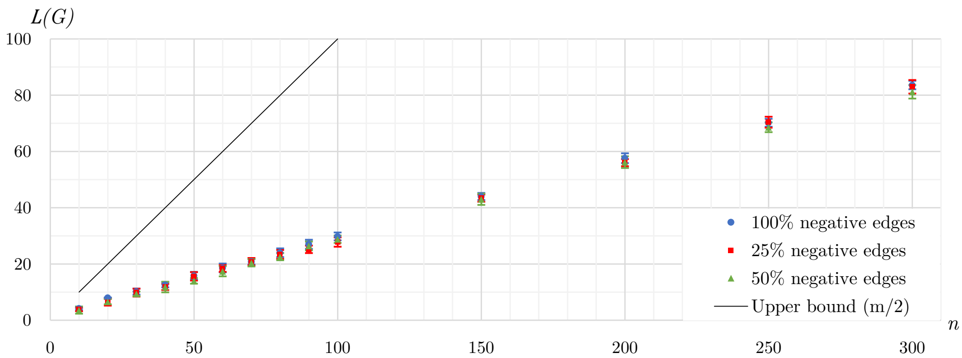


Figure 2: (colour online) The frustration index in random 4-regular networks of different orders n and decreasing densities $4/n - 1$

An observation to derive from Figure 2 is the similar frustration index values obtained for networks of the same sizes, even if they have different percentages of negative edges. It can be concluded that starting with an all-positive graph, making the first quarter of graph edges negative increases the frustration index much more than making further edges negative. Another observation is that the gap between the frustration index values and the theoretical upper bound increases with increasing n. Finding values of the frustration index closer to the theoretical upper bound may require investigating denser graphs with specific structures that we refer to as *maximally unbalanced graphs*.

6.3 Complete graphs with maximal imbalance

There have been a few studies of maximally unbalanced graphs and maximum frustration count values [7,40]. However, due to a lack of symmetry in balance theory, frustration maximising assignment of signs to a given graph is not known. In this section, two previously suggested families of specially structured complete graphs are investigated to help find an upper bound on the frustration index that is tighter than $m/2$.

The first family of specially structured graphs we consider is referred to as maximally unbalanced graphs by [7]. These graphs, denoted by K_n^b, are comprised of one cycle of n positive edges with the remaining pairs of nodes connected by negative edges forming a complete graph. The adjacency matrix can be defined as $\mathbf{A}(K_n^b) = 2C_n - K_n$ in which C_n is a cycle graph of order n and K_n is a complete graph of the same order. Based on the structure of the optimal solution given by (18) for K_n^b, we may calculate exact closed-form formulae for $L(K_n^b)$ as follows in (22).

$$L(K_n^b) = \begin{cases} (n^2 - 6n + 16)/4 & \text{if } n \text{ is even} \\ (n^2 - 6n + 17)/4 & \text{if } n \text{ is odd} \end{cases} \tag{22}$$

The second family of specially structured graphs to analyse includes all-negative complete graphs denoted by K_n^c. Associating the fraction of negative edges with lack of balance, one may expect K_n^c to have a high frustration. A simple observation of optimal solutions of model Eq. (18) for K_n^c reveals a structure in the solution that can be used to derive an exact closed-form formula for frustration index in all-negative complete graphs $L(K_n^c)$.

$$L(K_n^c) = \begin{cases} (n^2 - 2n)/4 & \text{if } n \text{ is even} \\ (n^2 - 2n + 1)/4 & \text{if } n \text{ is odd} \end{cases} \tag{23}$$

Eq. (23) confirms the upper bound $L(G) \leq \left\lfloor (n-1)^2/4 \right\rfloor$ Tomescu provided for the frustration index [39].

A comparison of $L(K_n^c)$ and $L(K_n^b)$ with $m/2 = (n^2 - n)/4$ reveals the gaps between the frustration index of such graphs and the proposed upper bound. This gap for even n equals $n/4$ and $(5n - 16)/4$ respectively for K_n^c and K_n^b. This supports the previous discussions about looseness of $m/2$ as an upper bound for frustration index. Assuming K_n^c to be "the family of maximally unbalanced graphs", $m/2 - n/4$ would be a tighter upper bound for the frustration index.

6.4 Performance of the binary linear models

In this section we discuss the time performance of the branch and bound algorithms for solving the binary linear models.

In order to compare the performance of the three linear models, we generate 10 decent-sized Erdős-Rényi random test cases with various densities and percentages of negative edges. As it can be observed in Table 2, AND has the smallest root node objective and ABS features the best root node objective amongst the models. The superiority of ABS is also confirmed in the solve time of the 10 test cases, while XOR is the slowest model. Keeping in mind that the number of B&B nodes depends on the heuristics and should not be considered as a definite criterion, it also indicates ABS as the best model in 8 out of 10 test cases while AND explores less B&B nodes in test cases 1 and 2.

Having gained an understanding of frustration in random signed graphs and the performance of the models, we continue to investigating frustration in small signed networks inferred from the literature.

7 Numerical results in real signed networks

In this section, the frustration index is computed in various real networks by solving the ABS model (19) using Gurobi Python interface and a desktop computer with an Intel Corei5 4670 @ 3.40 GHz and 8.00 GB of RAM running 64-bit Microsoft Windows 7.

There are well-studied signed social network datasets representing communities with positive and negative interactions and preferences. Read's dataset for New Guinean highland tribes [43] and Sampson's dataset for monastery interactions [44] are denoted by G1 and G2. We also use graphs inferred from datasets of students' choice and rejection denoted by G3 and G4 [45, 46]. A further explanation on the details of inferring signed graphs from the choice and rejection data can be found in [8]. Moreover, a larger signed network, denoted by G5, is inferred by [47] through implementing a stochastic degree sequence model on Fowler's data on Senate bill cosponsorship [48].

Besides the signed social network datasets, large scale biological networks can be analysed as signed graphs. There are four signed biological networks analysed by [49] and [50]. G6 is a network representing Epidermal growth factor receptor pathway [51]. G7 represents the molecular interaction map of a macrophage [52]. We also investigate two gene regulatory networks, related to two organisms: a eukaryote (the yeast Saccharomyces cerevisiae) and a bacterium (Escherichia coli). G8 Represents the gene regulatory network of Saccharomyces cerevisiae [53]. The largest biological network we consider is G9 which is related to the gene regulatory network of the Escherichia coli [54]. For more details on the previous records of the four biological datasets, one may refer to [50].

The results are shown in Table 3 where the average and standard deviation of the frustration index in 100 reshuffled graphs, denoted by $L(G_r)$ and SD, are also provided for comparison. Reshuffling the signs on the edges 100 times, we obtain two parameters of frustration distribution for the fixed un-

Table 2: Comparison of the the three binary linear models based on three performance measures

Test case	n	m	m^-	Density	m^-/m	Z^*	Root node objective			Number of B&B nodes			Solve time (s)		
							AND	XOR	ABS	AND	XOR	ABS	AND	XOR	ABS
1	65	570	395	0.27	0.69	189	-8.5	7.5	13.0	5133	24428	8316	65.4	714.3	22.2
2	68	500	410	0.22	0.82	162	-25.5	11.0	12.0	4105	48949	5240	27.3	462.8	8.8
3	80	550	330	0.17	0.60	170	-12.0	7.0	11.5	11652	34683	3216	153.3	1012.2	37.1
4	50	520	385	0.42	0.74	185	-82.5	12.5	13.5	901	25199	429	22.4	542.3	5.1
5	53	560	240	0.41	0.43	193	-130.0	14.5	15.5	292	25439	216	13.5	465.1	3.4
6	50	510	335	0.42	0.66	178	-24.0	7.0	15.0	573	32575	212	13.8	315.2	2.7
7	59	590	590	0.34	1.00	213	-52.0	12.5	15.0	1831	43630	460	46.0	966.3	11.9
8	56	600	110	0.39	0.18	110	-78.5	9.0	15.5	0	77	0	0.4	2.7	0.1
9	71	500	190	0.20	0.38	155	-7.5	8.0	11.0	6305	46713	2302	77.7	504.8	26.9
10	80	550	450	0.17	0.82	173	-7.5	8.0	11.5	12384	28762	2689	138.0	664.6	34.1

derlying structure. The randomisation process allocates signs to edges based on random permutations while preserving the unsigned graph structure.

Table 3: The frustration index in various signed networks

Graph	n	m	m^-	$L(G)$	$L(G_r) \pm$ SD	Z score
G1	16	58	29	7	14.80 ± 1.25	-6.25
G2	18	49	12	5	10.02 ± 1.22	-4.10
G3	17	40	17	4	8.02 ± 0.88	-4.55
G4	17	36	16	6	7.04 ± 1.00	-1.04
G5	100	2461	1047	331	973.83 ± 9.30	-69.13
G6	329	779	264	193	148.82 ± 5.11	8.65
G7	678	1425	478	332	253.16 ± 6.48	12.16
G8	690	1080	220	41	114.90 ± 5.52	-13.39
G9	1461	3212	1336	371	651.58 ± 6.92	-40.55

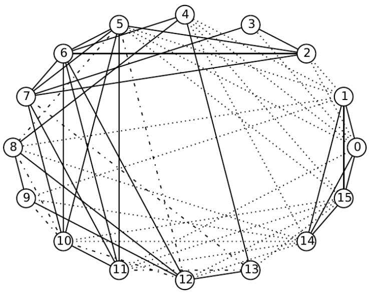
Although the signed networks G1 – G9 are not totally balanced, the relatively small values of $L(G)$ suggest low level of frustration in some of the the networks. G1 – G5 and G8 – G9 exhibit a level of frustration lower than expected by chance $L(G) < L(G_r)$.

Figure 3 shows how the small signed networks G1 – G4 can be made balanced by negating (or removing) the edges on the deletion minimal sets. Dotted lines represent negative edges, solid lines represent positive edges and edges in the deletion minimal set are indicated by dotdash lines regardless of their original signs.

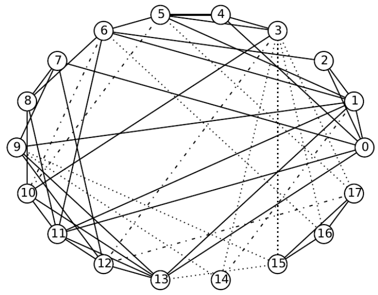
In order to be more precise, we have implemented a very basic statistical analysis using Z scores $Z = (L(G) - L(G_r))/\mathrm{SD}$. Z scores, provided in the right column of the Table 3, show how far the balance is with regards to balance distribution of the underlying structure. Negative values of Z score can be interpreted as lower level of frustration than the random expectation. Z score values also represent the significance when compared to the standard range of $(-3, 3)$.

The significance and level of partial balance are very high for G5 and G9, high for G1, G2, G3 and G8, and low for G4. The level of partial balance is very low for G6 and G7 which means that they are more frustrated than expected by chance.

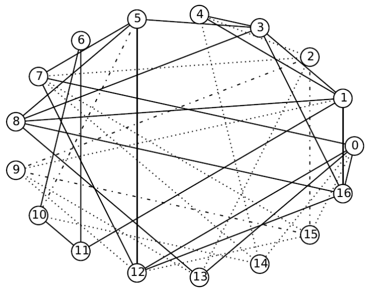
As the signed graph frustration problem has not been solved by exact methods, we compare the quality and solve time of our exact algorithm with that of recent heuristics and approximations implemented on the datasets. The capability limits of recent heuristics developed for the signed graph frustration index is discussed in Section 2. Various performance measures for the ABS model (19) solving G1 – G9 is illustrated in Table 4. The signed networks G1 – G4 are so small that a very basic binary linear formulation of the problem would solve them in a reasonable time. However for G5 – G9,



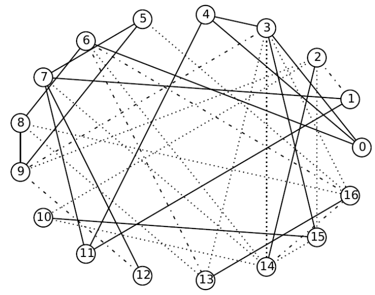
(a) Highland tribes network (G1), a signed network of 16 tribes of the Eastern Central Highlands of New Guinea [43]. Deletion minimal set comprises of 7 negative edges



(b) Monastery interactions network (G2) of 18 New England novitiates inferred from the integration of all positive and negative relationships [44]. Deletion minimal set comprises of 2 positive and 3 negative edges



(c) Fraternity preferences network (G3) of 17 boys living in a pseudo-dormitory inferred from ranking data of the last week in [45]. Deletion minimal set comprises of 4 negative edges



(d) College preferences network (G4) of 17 girls at an Eastern college inferred from ranking data of house B in [46]. Deletion minimal set comprises of 3 positive and 3 negative edges

Figure 3: The deletion minimal sets represented by dotdash lines for four small signed networks inferred from the sociology datasets

the speed-up techniques discussed are more useful to restrict the feasible space of more than 1000 binary variables.

Hüffner, Betzler, and Niedermeier have previously investigated frustration in G6 – G8 suggesting a data reduction scheme and an approximation algorithm [30]. Their suggested data reduction algorithm takes more than 5 hours for G8, more than 15 hours for G6, and more than 1 day for G7. Their approximation method provides $196 \leq L(G6) \leq 219$ and $L(G6) = 210$, both of which are incorrect. While their bound and exact solution for G8 is correct, they have also reported the exact value of $L(G7) = 374$ which is a sub-optimal solution.

Iacono et al. have investigated frustration in G6 – G9 [9]. Their heuristic algorithm provides upper and lower bounds for G6 – G9 with 96.37%,

Table 4: Performance measures for the ABS model (19) in solving G1 – G9

Graph	$L(G)$	Root node objective	Number of B&B nodes	Solve time (s)
G1	7	5.0	0	0.02
G2	5	4.0	0	0.02
G3	4	3.0	0	0.01
G4	6	3.5	0	0.01
G5	331	36.5	7	2.33
G6	193	15.5	2	0.49
G7	332	14.0	0	0.50
G8	41	11.5	0	0.29
G9	371	127.5	27	1.25

90.96%, 100%, and 98.38% ratio of lower to upper bound. Regarding solve time, they have only mentioned their heuristic requires a fairly limited amount of time (a few minutes on an ordinary PC).

Table 5 provides more comprehensive results on the solution quality and the solve time of the three linear models against the literature.

Table 5: Comparison of the solution quality and solve time between the linear models and against the literature

Graph		Hüffner et al. [30]	Iacono et al. [9]	AND	XOR	ABS
Quality	G6	[196, 219], 210	[186, 193]	193	193	193
	G7	[218,383], 374	[302, 332]	332	332	332
	G8	[0, 43], 41	41	41	41	41
	G9	Not tested	[365, 371]	371	371	371
Time	G6	15 hours	A few minutes	2.72 s	6.24 s	0.49 s
	G7	1 day	A few minutes	2.38 s	19.99 s	0.50 s
	G8	5 hours	A few minutes	0.91 s	0.74 s	0.29 s
	G9	Not tested	A few minutes	44.09 s	1.57 s	1.25 s

While data reductions schemes [30] take up to 1 day for these datasets and heuristic algorithms [9] only provide bounds with up to 9% gap from optimality, our ABS model solves the 9 datasets to optimality in a few seconds.

8 Temporal network frustration index

In this section, we analyse the frustration index in a political network of international relations over time. The Correlates of War dataset contains 51 time windows of a temporal network representing signed international relations among countries starting with 1946-1949 time window and ending with 1996-1999 time window [29].

In the first time window of the temporal network $n = 64$, $m = 362$ and $m^- = 42$. These numbers change in each time window as a result of the changes not only in the relations, but change of the nodes. In the last time window, $n = 151$, $m = 1247$ and $m^- = 147$. Figure 4 demonstrates the frustration index in the Correlates of War dataset. The solve times of the ABS model for each time window of this dataset is ≤ 0.18 seconds.

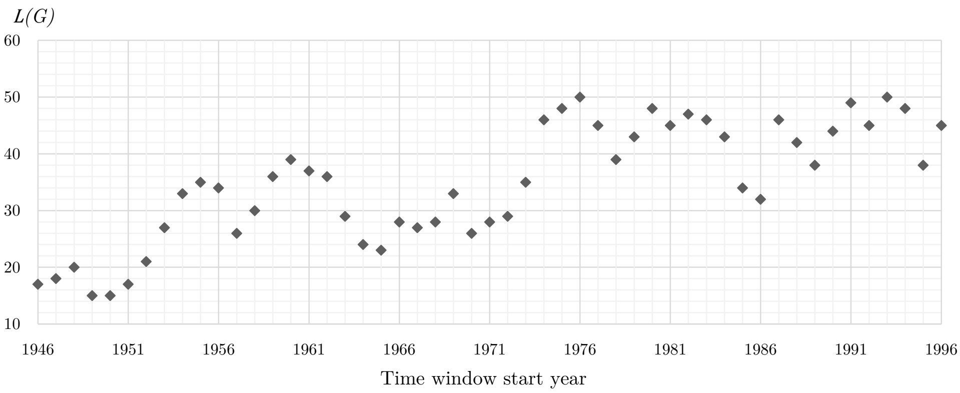


Figure 4: The frustration index of the Correlates of War dataset over time

Doreian and Mrvar have also analysed the Correlates of War dataset [29] using the frustration index [13] as a measure of balance. They claim that the frustration index is computable in polynomial time. Their argument must be evaluated while maintaining a certain level of healthy skepticism as their reported frustration index values using blockmodeling in Pajek is suboptimal for all 51 time windows of the Correlates of War dataset.

Bearing in mind that the size and order changes in each time window of the temporal network, we use the normalised frustration index, $F(G) = 1 - 2L(G)/m$, in order to investigate the partial balance over time. Figure 5 demonstrates the normalised frustration index in the Correlates of War dataset. This measure shows that the signed international relation network is close to a state of structural balance.

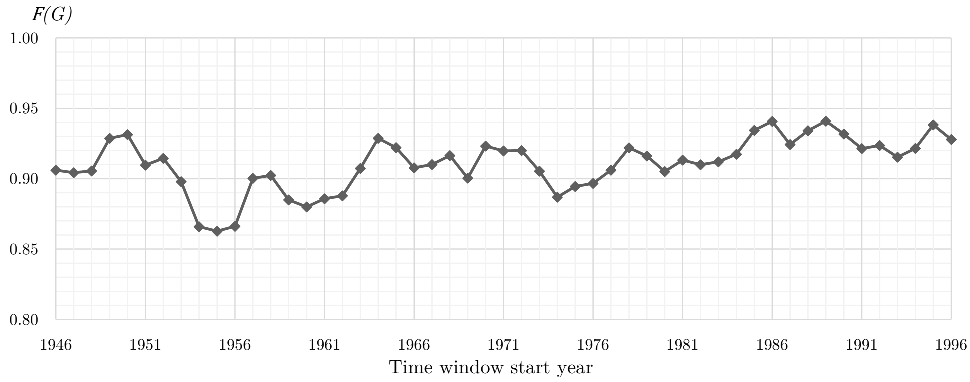


Figure 5: The partial balance in the Correlates of War dataset over time

9 Conclusion and future direction

Although this study focuses on frustration index as a measure of balance in signed networks, the findings may well have a bearing on the applications of frustration index in other disciplines discussed in Section 2. The present study suggested a novel method for computing a standard measures of partial balance that can be used for analysing network dynamics. It contributes additional evidence that suggests signed social networks and biological gene regulatory networks exhibit a relatively low level of frustration (relatively high partial balance). The findings for maximally unbalanced families of complete graphs suggest the looseness the previously suggested upper bounds even for graphs generated deliberately unbalanced.

This study have a number of important implications for future investigation. The optimisation model introduced can make network dynamics models more consistent with the theory [55]. To be more specific, many research studies on network dynamics use the number of balanced triads in the network as a criterion for transitioning towards balance using sequential sign change models. These models may result in stable states that are not totally balanced like jammed states and glassy states [56]. This contradicts not only the instability of unbalanced states, but the fundamental assumption that networks gradually moves towards balance. Deploying decrease in the frustration index as the criterion, the above-mentioned states can be avoided resulting in a realistic simulation of signed network dynamics that is consistent with the underlying theory and the assumptions.

References

[1] F. Heider, "Social perception and phenomenal causality." *Psychological review*, vol. 51, no. 6, pp. 358–378, 1944.

[2] D. Cartwright and F. Harary, "Structural balance: a generalization of Heider's theory." *Psychological review*, vol. 63, no. 5, pp. 277–293, 1956. [Online]. Available: http://psycnet.apa.org/journals/rev/63/5/277/

[3] F. Harary, "On the measurement of structural balance," *Behavioral Science*, vol. 4, no. 4, pp. 316–323, Oct. 1959. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/bs.3830040405/abstract

[4] R. Z. Norman and F. S. Roberts, "A derivation of a measure of relative balance for social structures and a characterization of extensive ratio systems," *Journal of mathematical psychology*, vol. 9, no. 1, pp. 66–91, 1972. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0022249672900065

[5] E. Terzi and M. Winkler, "A spectral algorithm for computing social balance," in *Algorithms and models for the web graph.* Springer, 2011, pp. 1–13. [Online]. Available: http://link.springer.com/chapter/10.1007/978-3-642-21286-4_1

[6] J. Kunegis, "Applications of Structural Balance in Signed Social Networks," *arXiv:1402.6865 [physics]*, Feb. 2014, arXiv: 1402.6865. [Online]. Available: http://arxiv.org/abs/1402.6865

[7] E. Estrada and M. Benzi, "Walk-based measure of balance in signed networks: Detecting lack of balance in social networks," *Physical Review E*, vol. 90, no. 4, pp. 1–10, 2014. [Online]. Available: http://journals.aps.org/pre/abstract/10.1103/PhysRevE.90.042802

[8] S. Aref and M. C. Wilson, "Measuring partial balance in signed networks," *arXiv preprint arXiv:1509.04037*, 2016.

[9] G. Iacono, F. Ramezani, N. Soranzo, and C. Altafini, "Determining the distance to monotonicity of a biological network: a graph-theoretical approach," *Systems Biology, IET*, vol. 4, no. 3, pp. 223–235, 2010. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5470321

[10] F. Harary, M.-H. Lim, and D. C. Wunsch, "Signed graphs for portfolio analysis in risk management," *IMA Journal of Management Mathematics*, vol. 13, no. 3, pp. 201–210, Jan. 2002. [Online]. Available: http://imaman.oxfordjournals.org/content/13/3/201

[11] D. Cimasoni, "Computing the writhe of a knot," *Journal of Knot Theory and Its Ramifications*, vol. 10, no. 03, pp. 387–395, May 2001. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S0218216501000913

[12] P. W. Kasteleyn, "Dimer Statistics and Phase Transitions," *Journal of Mathematical Physics*, vol. 4, no. 2, pp. 287–293, Feb. 1963. [Online]. Available: http://scitation.aip.org/content/aip/journal/jmp/4/2/10.1063/1.1703953

[13] P. Doreian and A. Mrvar, "Structural Balance and Signed International Relations," *Journal of Social Structure*, vol. 16, pp. 1–49, 2015.

[14] T. Doslic and D. Vukicevic, "Computing the bipartite edge frustration of fullerene graphs," *Discrete Applied Mathematics*, vol. 155, no. 10, pp. 1294–1301, May 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0166218X06005129

[15] C. Chiang, A. B. Kahng, S. Sinha, X. Xu, and A. Z. Zelikovsky, "Fast and Efficient Bright-Field AAPSM Conflict Detection and Correction," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 1, pp. 115–126, Jan. 2007.

[16] T. Zaslavsky, "Balance and clustering in signed graphs," *Slides from lectures at the CR Rao Advanced Institute of Mathematics, Statistics and Computer Science, Univ. of Hyderabad, India*, vol. 26, 2010. [Online]. Available: http://vulcan.math.binghamton.edu/zaslav/Tpapers/balance-clusterability.slides.20100726.pdf

[17] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin-glass," *Phys. Rev. Lett.*, vol. 35, pp. 1792–1796, Dec 1975. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.35.1792

[18] F. Barahona, "On the computational complexity of ising spin glass models," *Journal of Physics A: Mathematical and General*, vol. 15, no. 10, p. 3241, 1982. [Online]. Available: http://stacks.iop.org/0305-4470/15/i=10/a=028

[19] F. Hadlock, "Finding a maximum cut of a planar graph in polynomial time," *SIAM Journal on Computing*, vol. 4, no. 3, pp. 221–225, 1975. [Online]. Available: http://dx.doi.org/10.1137/0204019

[20] M. X. Goemans and D. P. Williamson, "Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming," *J. ACM*, vol. 42, no. 6, pp. 1115–1145, Nov. 1995. [Online]. Available: http://doi.acm.org/10.1145/227683.227684

[21] O. Katai and S. Iwai, "Studies on the balancing, the minimal balancing, and the minimum balancing processes for social groups with planar and nonplanar graph structures," *Journal of Mathematical Psychology*, vol. 18, no. 2, pp. 140–176, 1978.

[22] J. A. Davis, "Clustering and structural balance in graphs," *Human Relations*, vol. 20, no. 2, pp. 181–187, 1967. [Online]. Available: http://hum.sagepub.com/content/20/2/181.short

[23] I. Giotis and V. Guruswami, "Correlation clustering with a fixed number of clusters," in *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithm*, ser. SODA '06. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2006, pp. 1167–1176. [Online]. Available: http://dl.acm.org/citation.cfm?id=1109557.1109686

[24] M. Brusco and D. Steinley, "K-balance partitioning: An exact method with applications to generalized structural balance and other psychological contexts," *Psychological Methods*, vol. 15, no. 2, pp. 145–157, 2010, 00010.

[25] R. Figueiredo and G. Moura, "Mixed integer programming formulations for clustering problems related to structural balance," *Social Networks*, vol. 35, no. 4, pp. 639–651, 2013.

[26] L. Drummond, R. Figueiredo, Y. Frota, and M. Levorato, "Efficient solution of the correlation clustering problem: An application to structural balance," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 2013, pp. 674–683.

[27] M. Levorato, L. Drummond, Y. Frota, and R. Figueiredo, "An ils algorithm to evaluate structural balance in signed social networks," in *Proceedings of the 30th Annual ACM Symposium on Applied Computing*. ACM, 2015, pp. 1117–1122.

[28] E. D. Demaine, D. Emanuel, A. Fiat, and N. Immorlica, "Correlation clustering in general weighted graphs," *Theoretical Computer Science*, vol. 361, no. 2, pp. 172–187, 2006.

[29] J. Pevehouse, T. Nordstrom, and K. Warnke, "The correlates of war 2 international governmental organizations data version 2.0," *Conflict Management and Peace Science*, vol. 21, no. 2, pp. 101–119, 2004.

[30] F. Hüffner, N. Betzler, and R. Niedermeier, "Separator-based data reduction for signed graph balancing," *Journal of combinatorial optimization*, vol. 20, no. 4, pp. 335–360, 2010. [Online]. Available: http://link.springer.com/article/10.1007/s10878-009-9212-2

[31] G. Facchetti, G. Iacono, and C. Altafini, "Computing global structural balance in large-scale signed social networks," *Proceedings of the National Academy of Sciences*, vol. 108, no. 52, pp. 20 953–20 958, 2011.

[32] P. Esmailian, S. E. Abtahi, and M. Jalili, "Mesoscopic analysis of online social networks: The role of negative ties," *Physical Review E*, vol. 90, no. 4, p. 042817, 2014. [Online]. Available: http://journals.aps.org/pre/abstract/10.1103/PhysRevE.90.042817

[33] L. Ma, M. Gong, H. Du, B. Shen, and L. Jiao, "A memetic algorithm for computing and transforming structural balance in signed networks," *Knowledge-Based Systems*, vol. 85, pp. 196 – 209, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705115001811

[34] L. Ma, M. Gong, J. Yan, F. Yuan, and H. Du, "A decomposition-based multi-objective optimization for simultaneous balance computation and transformation in signed networks," *Information Sciences*, vol. 378, pp. 144–160, Feb. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0020025516313883

[35] G. Facchetti, G. Iacono, and C. Altafini, "Exploring the low-energy landscape of large-scale signed social networks," *Physical Review E*, vol. 86, no. 3, p. 036116, 2012.

[36] P. Esmailian and M. Jalili, "Community detection in signed networks: The role of negative ties in different scales," *Scientific reports*, vol. 5, 2015.

[37] I. Mendonça, R. Figueiredo, V. Labatut, and P. Michelon, "Relevance of negative links in graph partitioning: A case study using votes from the european parliament," in *Network Intelligence Conference (ENIC), 2015 Second European*. IEEE, 2015, pp. 122–129.

[38] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. De Luca, and S. Albayrak, "Spectral analysis of signed graphs for clustering, prediction and visualization." in *SDM*, vol. 10. SIAM, 2010, pp. 559–570. [Online]. Available: http://epubs.siam.org/doi/pdf/10.1137/1.9781611972801.49

[39] I. Tomescu, "Note sur une caractérisation des graphes dont le degré de déséquilibre est maximal," *Mathématiques et sciences humaines*, vol. 42, pp. 37–40, 1973. [Online]. Available: http://archive.numdam.org/article/MSH_1973__42__37_0.pdf

[40] J. Akiyama, D. Avis, V. Chvàtal, and H. Era, "Balancing signed graphs," *Discrete Applied Mathematics*, vol. 3, no. 4, pp. 227–233,

Nov. 1981. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0166218X81900019

[41] G. Facchetti, G. Iacono, and C. Altafini, "Computing global structural balance in large-scale signed social networks," *Proceedings of the National Academy of Sciences*, vol. 108, no. 52, pp. 20 953–20 958, Dec. 2011. [Online]. Available: http://www.pnas.org/content/108/52/20953

[42] E. Klotz and A. M. Newman, "Practical guidelines for solving difficult mixed integer linear programs," *Surveys in Operations Research and Management Science*, vol. 18, no. 12, pp. 18 – 32, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1876735413000020

[43] K. E. Read, "Cultures of the central highlands, New Guinea," *Southwestern Journal of Anthropology*, pp. 1–43, 1954. [Online]. Available: http://www.jstor.org/stable/3629074

[44] S. F. Sampson, "A Novitiate in a Period of Change," *An Experimental and Case Study of Social Relationships (PhD thesis). Cornell University, Ithaca*, 1968.

[45] T. Newcomb, *The Acquaintance Process. New York: Holt, Rinehart and Winston.(1966)." The General Nature of Peer Group Influence", pps. 2-16 in College Peer Groups, edited by TM Newcomb and EK Wilson.* Chicago: Aldine Publishing Co, 1961.

[46] T. B. Lemann and R. L. Solomon, "Group characteristics as revealed in sociometric patterns and personality ratings," *Sociometry*, vol. 15, pp. 7–90, 1952. [Online]. Available: http://www.jstor.org/stable/2785447

[47] Z. Neal, "The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors," *Social Networks*, vol. 39, pp. 84–97, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378873314000343

[48] J. H. Fowler, "Legislative cosponsorship networks in the US House and Senate," *Social Networks*, vol. 28, no. 4, pp. 454–465, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378873305000730

[49] B. DasGupta, G. A. Enciso, E. Sontag, and Y. Zhang, "Algorithmic and complexity results for decompositions of biological networks into monotone subsystems," *Biosystems*, vol. 90, no. 1, pp. 161–178, 2007.

[50] G. Iacono, F. Ramezani, N. Soranzo, and C. Altafini, "Determining the distance to monotonicity of a biological network: a graph-theoretical approach," *IET Systems Biology*, vol. 4, no. 3, pp. 223–235, 2010.

[51] K. Oda, Y. Matsuoka, A. Funahashi, and H. Kitano, "A comprehensive pathway map of epidermal growth factor receptor signaling," *Molecular systems biology*, vol. 1, no. 1, 2005.

[52] K. Oda, T. Kimura, Y. Matsuoka, A. Funahashi, M. Muramatsu, and H. Kitano, "Molecular interaction map of a macrophage," *AfCS Research Reports*, vol. 2, no. 14, pp. 1–12, 2004.

[53] M. C. Costanzo, M. E. Crawford, J. E. Hirschman, J. E. Kranz, P. Olsen, L. S. Robertson, M. S. Skrzypek, B. R. Braun, K. L. Hopkins, P. Kondu, C. Lengieza, J. E. Lew-Smith, M. Tillberg, and J. I. Garrels, "Ypd, pombepd and wormpd: model organism volumes of the bioknowledge library, an integrated resource for protein information," *Nucleic Acids Research*, vol. 29, no. 1, pp. 75–79, 2001. [Online]. Available: http://nar.oxfordjournals.org/content/29/1/75.abstract

[54] H. Salgado, S. Gama-Castro, M. Peralta-Gil, E. Diaz-Peredo, F. Sánchez-Solano, A. Santos-Zavaleta, I. Martinez-Flores, V. Jiménez-Jacinto, C. Bonavides-Martinez, J. Segura-Salazar *et al.*, "Regulondb (version 5.0): Escherichia coli k-12 transcriptional regulatory network, operon organization, and growth conditions," *Nucleic acids research*, vol. 34, no. suppl 1, pp. D394–D397, 2006.

[55] T. Antal, P. L. Krapivsky, and S. Redner, "Dynamics of social balance on networks," *Physical Review E*, vol. 72, no. 3, p. 036121, 2005. [Online]. Available: http://journals.aps.org/pre/abstract/10.1103/PhysRevE.72.036121

[56] S. A. Marvel, S. H. Strogatz, and J. M. Kleinberg, "Energy landscape of social balance," *Physical review letters*, vol. 103, no. 19, p. 198701, 2009. [Online]. Available: http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.103.198701