A graph modelling to measure the frustration index in signed networks

Germina Augusthy Kizhakekunnel
Department of Mathematics
Central University of Kerala
Kasargod, Kerala
India
germinaka@cukerala.ac.in

Abstract. Computing the frustration index of a signed graph is a key step towards solving problems in many fields including social networks, political science, physics, chemistry, and biology. In social networks the frustration index determines network distance from a state of structural balance. The focus of this paper is to provide insight into computing the frustration index and show that exact values of the frustration index can be computed using optimisation models namely, 0/1 Binary Linear Model.

Keywords: frustration index, line index of balance, signed graph, integer programming, optimisation, balance theory.

1. Introduction

Heider published his important article on the psychology of cognitive organization in 1946. Since then, theories of cognitive consistency or psychological balance have proliferated remarkably. The rise of 'balance theories' to such popularity is explainable in part by their theoretical elegance and simplicity, but also by their apparent applicability to a wide range of problems in sociology and social psychology. Reference group relations, strain toward consistency in culture, persuasive communication, socialization of self-attitudes, friendship formation, the formation of group norms and value consensus, group or clique formation, group reactions to deviant behavior, the effect of initiation rites on group solidarity, reactions to status inconsistency, even some aspects of international relations, all can be given psychologically based interpretations with the aid of some form of balance theory.

Psychological principles with such broad explanatory potential clearly warrant the critical attention of sociologists. Social relations are complex and subtle (see, [16, 12]), reflecting interactions that could be either positive such as trust, agreement, approval; or negative such as distrust, disagreement, opposition. Yet until recently a number of authors began to investigate signed relations by examining the links in emerging signed social networks. Tom Coleman et.al. [16] examined the applicability of social psychology theories at multiple signed networks. F. Harrary et. al [3] further applied classification approaches to predict the sign of social relationships. Many researchers focus on discovering the interplay between relations and behavior. The nontrivial relationship between behavioral correlation and the sign of social relations motivated researchers to leverage both social and behavioral evidences when predicting signed social ties. Hence, study has been done to design models to capture the interplay between users- behavior of decision making and social interaction in the context of social networks. The problem of predicting signed social ties, such as trust and distrust, based on the acquaintance relationships in social networks thereby to determine, whether a link corresponds to a trustworthy friend or rather an enemy. The frustration index is one of the key measure for analysing signed networks.

A better understanding of balance theory and computation of frustration index can be done using signed graph models. This paper present a graph model that infer signed ties by capturing the interplay between social relations and users- behavior of decision making, and thereby compute whether the given system is balanced or not. For that we need the terms and concepts in graphs and signed graphs.

2. Graph theoretic concepts

All graphs in this paper are simple and finite. Let $G = (V, E)$ be a simple graph of order n and size m with vertex set $V = \{v_1, v_2, \ldots, v_n\}$ and edge set $E = \{e_1, e_2, \ldots, e_m\}$. For the terms and definitions not mentioned in the paper one may refer [13].

Signed graphs (also called sigraphs) are much studied in literature because of their extensive use in modeling a variety of socio-psychological processes and also because of their interesting connections with many classical mathematical systems (see, [2, 8, 18, 19, 20]).

Formally, a signed graph is defined as

Definition 2.1. *A signed graph is an ordered pair* $\Sigma = (G, \sigma)$ *where* $G = (V, E)$ *is a graph called the underlying graph of* Σ *and* $\sigma : E \to \{+1, -1\}$*, is a function called a* signature or signing . $E^+(\Sigma)$ *denotes the set of all edges of* G *that are mapped by* σ *to the element* $+1$ *and* $E^-(\Sigma) = E - E^+(\Sigma)$*. The elements of* $E^+(\Sigma)$ *are called positive edges and those of* $E^-(\Sigma)$ *are called negative edges of* Σ*.*

A signed graph is *all-positive* (respectively, *all-negative*) if all of its edges are positive (negative); further, it is said to be *homogeneous* if it is either all-positive or all-negative and *heterogeneous* otherwise. The sign of a cycle in a signed graph is the product of the signs of its edges. Thus, a cycle is positive if and only if it contains even number of negative edges. A signed graph Σ is said to be *balanced* or *cycle balanced* if all of its cycles are positive. Two signed graphs, $\Sigma_1 = (G_1, \sigma_1)$ and $\Sigma_2 = (G_2, \sigma_2)$, are *isomorphic* if there is a graph isomorphism $f : G_1 \to G_2$ that preserves signs of the edges. If $\theta : V \to \{+1, -1\}$ is *switching* *function*, then *switching* of the signed graph $\Sigma = (G, \sigma)$ by θ means changing σ to σ^θ defined by:

$$\sigma^\theta(uv) = \theta(u)\sigma(uv)\theta(v).$$

The switched graph denoted by Σ^θ, is the signed graph $\Sigma^\theta = (G, \sigma^\theta)$. We call two signed graphs $\Sigma_1 = (G, \sigma_1)$ and $\Sigma_2 = (G, \sigma_2)$ to be *switching equivalent* and write $\Sigma_1 \sim \Sigma_2$, if there exists a switching function $\theta : V \to \{+1, -1\}$ such that $\Sigma_1 = \Sigma_2^\theta$. It can be seen that switching preserves many features of the two signed graphs including their eigenvalues [14] and balance.

In the classic friend-enemy interpretation of balance theory, a by-product of computing the frustration index is the partitioning of nodes into two internally solidary but mutually hostile groups.

Theorem 2.1 ([3])**.** *A signed graph is balanced if and only if, the vertex set V can be partitioned into two subgraphs, such that vertices within a subgraph are connected by a positive edge and vertices in separate subgraphs are connected by a negative edge.*

More generally, a signed graph Σ is said to be **clusterable** or **partitionable** if its vertex set can be partitioned into subsets (**clusters**), so that every positive edge joins the vertices within the same cluster and every negative edge joins the vertices in the different clusters. The question to be addressed is that, is it possible to partition vertices of a signed graph, so that every line that connects vertices that belong to the same cluster is positive and every line that connects two vertices that belong to different clusters is negative? If it is possible to partition vertices in this way, we call the signed graph partitionable or clusterable. However, it is a general rule that positive and negative relations among people tend towards balance. Hence, important are balanced signed graphs where vertices can be partitioned into 2 clusters.

2.1 Computation of frustration index in a signed network

Computing the frustration index of a signed graph is a key step towards solving problems in many fields including social networks, physics, material science, and biology. The frustration index determines the distance of a network from a state of total structural balance. 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 versus enmity or trust versus distrust between people. The term signed network embodies a multitude of concepts involving relationships characterizable 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 become a focus in network science. Structural balance theory was the first attempt to understand the sources of tensions and conflicts in groups of people with signed ties (see, [15]). According to balance theory, some structural configurations of people with signed ties lead to social tension and therefore are not balanced. Using graph-theoretic concepts, Cartwright and Harary identified cycles of the graph as the origins of tension, in particular cycles containing an odd number of negative edges [3]. By definition, signed graphs in which no such negative cycles are present satisfy the property of structural balance. The vertex set of a balanced signed network can be partitioned into $k \geq 2$ subsets such that each negative edge joins vertices belonging to different subsets [8]. For graphs that are not totally balanced, a distance from total balance called the *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 balance (see, [18, 19, 20]). Measures based on cycles [7], triangles, and closed walks [17] are not generally consistent and do not satisfy key axiomatic properties [12]. Among all the measures, a normalised version of the frustration index is shown to satisfy many basic axioms [12]. This measure provides a clear understanding of the transition to balance in terms of the number of edges to be modified to reduce the tension, as opposed to graph cycles that were first suggested as origins of tension in unbalanced networks.

2.2 Complexity of computing frustration index

Computing the frustration index is related to the well-known unsigned graph optimization problem edge-bipartization, which requires minimization of the number of edges whose deletion makes the graph bipartite. Given an instance of the latter problem, by declaring each edge to be negative we convert it to the problem of computing the frustration index. Since edge-bipartization is known to be NP-hard, so is computing the frustration index. In the converse direction there is a reduction of the frustration index problem to edge-bipartization which increases the number of edges by a factor of at most 2. Hence, the frustration index can be computed in polynomial time for planar graphs (see, [9]), which is equivalent to the ground state calculation of a two-dimensional spin glass model with no periodic boundary conditions and no magnetic field [16]. The classic graph optimization problem MAXCUT is also a special case of the frustration index problem, as can be seen by assigning all edges to be negative.

3. Frustration count

The number of edges incident to the node $i \in V$ represents the degree of node i and is denoted $d(i)$. By net degree of a vertex v in signed graph we mean the difference of number of positive and negative edges incident at the vertex v, denoted respectively as $d^+(v)$ (or $d^-(v)$), if the net degree is positive (or negative) at the the vertex v. We apply BFS Algorithm with the initial condition in choosing the root vertex in the given network graph as follows. Choose the root vertex say u such that net degree $d^{-}(u)$ is maximum. Visit the vertices that are in the first neighbourhood of the vertex u. Then visit the neighbous of all the vertices which are already visited, in the order in which they are visited. That is, represent the m undirected edges in G as ordered pairs of vertices $E = \{e_1, e_2, \ldots, e_m\} \subseteq \{(i,j), i < j, i, j \in V\}$ where a single edge e_k between nodes i and $j, i < j$, is denoted by $e_k = (i,j), i < j$. Let the number of positive and negative vertices in the nth level be respectively l_n^+ and l_n^-. BFS algorithm separates the edges into two distinct sets namely the set of tree edges and set of cycle edges (or back edges). Let us denote the set consisting of tree edges by T and the set consisting of cycle edges by C.

Let $V_i, 0 \leq i \leq n$ be the vertex set in level $l_i, 0 \leq i \leq n$. Then, $n+1$ is the number of levels fixing the root vertex u at the 0th level. Now consider the partition of the vertex set V, as $V_o = \bigcup_i V_i, i\ odd$ and $V_e = \bigcup_i V_i, i\ even$, where i denote the level number.

To compute the frustration index of the signed graph, we only need to concentrate on cycle edges, (that are either positive or negative forming cycles), as the sign of the cycles decide the balance of the signed graph. We prove the following Lemma first.

Lemma 3.1. *Consider the partition of the vertex set as described in BFS algorithm,* $V_o = \bigcup_i V_i, i\ odd$ *and* $V_e = \bigcup_i V_i, i\ even$*. Then, the number of negative cycle edges appear in the same level and the positive cycle edges in two consecutive levels, of the given signed graph decides the frustration index of the given signed graph.*

Proof. Consider the partition of the vertex set obtained by applying BFS algorithm, $V_o = \bigcup_i V_i, i\ odd$ and $V_e = \bigcup_i V_i, i\ even$.
Case 1. Negative cycle edges in C.

That is, negative cycle edges are in the same partitioned set V_o (or V_e), say at the ith level. We claim that these negative edges are the 'bad edges' or 'frustrated edges' in the signed graph. Now, if the negative edge in the same level l_i is such that the end vertices of the edges are incident to negative tree edges in level l_{i-1}, which again are incident with any father negative edges in level $l_{i-j}, 1 \leq j \leq i$, then we get a negative cycle $C_n, n \geq 3$. On the other hand, if the negative cycles edges appear in the different level, then the negative edges are such that one of the end vertices of the each of the negative edges are in V_e and other in V_o. Hence, by Harray's bipartition theorem, none of these edges are 'bad edges'. Let us call such edges as 'good' or 'satisfied' edges.
Case 2. Consider positive cycle edges in C.

If, the positive cycle edges appear in the different level, then the positive edges are such that one of the end vertices of the each of the positive edges are in V_e and other in V_o. Hence, by Harray's bipartition theorem, all of these edges are 'bad edges'. On the other hand, if the positive cycle edges with both their end vertices in the same partitioned set V_e or V_o, then by Harray's bipartition theorem, none of these positive edges are 'bad edges'.

Hence, deletion of negative cycle edges appear in the same level and the positive cycle edges in two consecutive levels, of the given signed graph, the given signed graph becomes balanced. □

Remark 3.1. Given a signed network, respectively a signed graph on n vertices, choosing a proper vertex as root vertex, and applying BFS Algorithm with this specified condition of choosing the root vertex, we get n^{n-2} spanning trees, with two sets of edges namely set with trees edges T and set C consisting of cycle edges. However, only cycle edges contribute to the frustration index of the signed graph. Hence, the problem of minimizing the frustration index depends on the choice of the spanning tree from the set of n^{n-2} spanning trees. Choose the spanning tree in such a way that it follows the following rules.
Case (i). Choose the spanning tree in which the number of negative edges with one end in V_0 and other end in V_e is maximum, and the spanning tree in which the number of negative cycle edges in same set V_o (and in V_e) are minimum.
Case (ii). Choose the spanning tree in which the number of positive edges with one end in V_0 and other end in V_e is minimum, and, the spanning tree in which the number of positive edges in same set V_o (and in V_e) is maximum.

Let us define these 'Good' and 'bad' edges based on colourings of the nodes. Colouring each node in V_o with black and the each node in V_e with white, a bad (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. On the other hand, a 'good' edge (i, j) is either a negative edge with different colours on the endpoints i, j or a positive edge with the same colours on the endpoints i, j. Hence, the problem of computing the frustration index reduces to optimizing the number of negative edges with same colours and the number of positive edges with different colours. That is, the problem of computing the frustration index of a given signed graph reduces to a linear programming problem.

4. Binary linear programming and frustration count

Now, $V_o \cup V_e = V$ defines a partition of V. By Colouring each node in V_0 with black and each node in V_e with white. One can view this partition in terms of coloring, namely, $V = X \cup V - X$, where $X \subseteq V$ a subset of vertices, with colors black or white. We call X a colouring set. Now, define the binary variable x_i to denote the colour of node $i \in V$ under this colouring set X as $x_i = 1$, if $i \in X$ (black node) and $x_i = 0$ if $i \in V - X$ (white node).

Hence, the frustration count of signed graph G under colouring X is $f_G(X) := \sum_{(i,j)\in E} f_{ij}(X)$ where $f_{ij}(X)$ is the frustration state of edge (i, j), given by

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

The frustration index of a signed graph Σ depends on the choice of the spanning tree satisfying the conditions Case(i) and Case (ii) of Remark 3.1. That is, the frustration index $L(G)$ of a graph G can be found by finding the spanning tree satisfying the conditions Case (i) and Case (ii) of Remark 3.1, such that the choice of spanning tree that minimizes the frustration count $f_G(X)$, which is obtained by solving the Equation (1).

Hence, the objective function Z of the Binary Linear Programming Model is to Minimize the frustration count. That is , the binary linear programming model is as follows.

Fix the objective function to minimize the frustration count. There arises two cases.

Case (i). For counting the frustration index for the positive cycle edges:

The frustration of a positive edge (i, j) be represented by $f_{ij} = x_i + x_j - 2(x_i \text{ and } x_j), \forall (i, j) \in E^+$, using the two binary variables x_i, x_j for the endpoint colours.

Case (ii). For counting the frustration index for the negative cycle edge:

For a negative edge $f_{ij} = 1 - (x_i + x_j - 2(x_i \text{ and } x_j)), \forall (i, j) \in E^-$.

The term x_i and x_j can be replaced by binary variables $x_{ij} = x_i$ and x_j for each edge (i, j) that take value 1 whenever $x_i = x_j = 1$ (both endpoints are coloured black) and 0 otherwise.

Hence, this model is a 0/1 linear programming model which computes the frustration index in the minimisation objective function as

$Minimize_{x_i \in V; (i,j) \in E} \quad Z = \sum_{(i,j) \in E^+} (x_i + x_j - 2x_{ij}) + \sum_{(i,j) \in E^-} 1 - ((x_i + x_j) - 2x_{ij})$

Subject to

$x_{ij} \leq x_i$, for all $(i, j) \in E^+$

$x_{ij} \leq x_j$, for all $(i, j) \in E^+$

$x_{ij} \geq x_i + x_j - 1$, for all $(i, j) \in E^-$

$x_i \in \{0, 1\}$, for all $i \in V$

$x_{ij} \in \{0, 1\}$, for all $(i, j) \in E$.

Remark 4.1. One may note that the dependencies between the x_{ij} and x_i, x_j values are taken into account in this constraints. This model has $n+m$ variables and $2m^+ + m^-$ constraints. The dependent variables are x_{ij}. Because of the constraints and the minimisation objective function, one may consider them as continuous variables in the unit interval, $x_{ij} \in [0, 1]$.

5. Conclusion

Solving the 0/1 linear model provides us with the minimum number of frustrated edges in the given signed graph. This number determines how many edges should be removed to make the network balanced. By applying BFS algorithm and selecting the spanning tree among the set of n^{n-2} spanning trees may be done by counting the number of negative edges in consecutive levels or by identifying and counting the vertices with maximum negative net degree. This process itself give a way to minimize the bad cycle edges in the signed graph so that the optimization process may converge faster.

One may be interested in generalizing the 2-colour minimum frustration count optimization problem to k- colour ($k \geq 2$). That is, the vertex set of the signed graph is partitioned onto k subsets ($k \geq 2$) such that each negative edges joins vertices belonging to disjoint subsets. The minimum number of frustrated edges for general k ($k \geq 3$) is referred as correlation clustering problem. Another generalization of 2 colour minimum frustration count optimization problem by allowing more than two colours, which is called multi-colour minimum frustration count optimization problem. Developing a new optimization model based on optimized multi-colour minimum frustration count is an interesting problem for further investigation.

Acknowledgement

The author would like to acknowledge SERB, Govt. of India for the financial support under MATRICS Scheme vide order NO. MTR/2017/000689.

References

[1] N. Alon, *Choice numbers of graphs: a probabilistic approach*, Combin. Probab. Comput., 1 (1992), 107-114.

[2] B.D. Acharya, *Spectral criterion for cycle balance in networks*, J. Graph Theory, 4 (1980), 1-11.

[3] Dorwin Cartwright, Frank Harary, *Structural balance: a generalization of Heider-s theory*, Psychological Review, 63 (1956), 277-293.

[4] P. Erdös, *On a combinatorial problem-I*, Nordisk. Mat. Tidskrift, 11 (1963), 5-10.

[5] P. Erdös, *On a combinatorial problem-II*, Acta Math. Hungar, 15 (1964), 445-447.

[6] P. Erdös, A.L. Rubin, H. Taylor, *Choosability in graphs*, Proc. West Coast Conf. on Combinatorics, Graph Theory and Computing, Congr. Numer., XXVI (1979), 125-157.

[7] F Barahona, *On the computational complexity of Ising spin glass models*, Journal of Physics A: Mathematical and General, 15 (1982), 3241-3253.

[8] F. Harary, R.Z. Norman, D. Cartwright, *Structural models: an introduction to the theory of directed graphs*, Wiley, New York 1965.

[9] Giuseppe Facchetti, Giovanni Iacono, and Claudio Altafini, *Computing global structural balance in large-scale signed social networks*, Proceedings of the National Academy of Sciences, 108 (2011), 20953-20958.

[10] Luerbio Faria, Sulamita Klein, Matej Stehlik, *Odd cycle transversals and independent sets in fullerene graphs*, SIAM Journal on Discrete Mathematics, 26 (2012), 1458-1469.

[11] Robert P. Abelson, Milton J. Rosenberg, *Symbolic psycho-logic: a model of attitudinal cognition*, Behavioral Science, 3 (1958), 1-13.

[12] Samin Aref, Andrew J Mason, Mark C Wilson, *An exact method for computing the frustration index in signed networks using binary programming*, arXiv:1611.09030, 2017.

[13] F. Harary, *Graph theory*, Narosa Publ. House, New Delhi, 2001.

[14] Osamu Katai, Sousuke 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, 18 (1978), 140-176.

[15] Rosa Figueiredo, Gisele Moura, *Mixed integer programming formulations for clustering problems related to structural balance*, Social Networks, 35 (2013), 639-651.

[16] Tom Coleman, James Saunderson, Anthony Wirth, *A local-search 2-approximation for 2-correlationclustering* In Dan Halperin and Kurt Mehlhorn, editors, European Symposium on Algorithms, pages (2008), 308-319, Berlin, Germany, Springer.

[17] Tomislav Doslic, *Bipartivity of fullerene graphs and fullerene stability*, Chemical Physics Letters, 412 (v), 336-340.

[18] T. Zaslavsky, *Signed graphs*, Discrete Appl. Math., 4 (1982), 47-74, Erratum, Discrete Appl. Math., 5 (1983), 248.

[19] T. Zaslavsky, *Matrices in the theory of signed simple graphs*, In: B.D. Acharya, G.O.H. Katona, and J. Nesetril, eds., *Advances in Discrete Mathematics and Its Applications* (Proc. Int. Conf. Discrete Math., Mysore, India, 2008). Ramanujan Math. Soc. Lecture Notes Ser. Math., to appear, 207-229.

[20] T. Zaslavsky, *A mathematical bibliography of signed and gain graphs and allied areas*, VII edition. Electronic J. Combinatorics, 8 (1998), Dynamic Surveys #DS8, 124 pp.

Accepted: March 23, 2021