Measuring Partial Balance in Signed Networks

Samin Aref
Department of Computer Science
University of Auckland
Auckland, Private Bag 92019, New Zealand
Email: sare618@aucklanduni.ac.nz

Mark C. Wilson
Department of Computer Science
University of Auckland
Email: mc.wilson@auckland.ac.nz

***Abstract*—Is the enemy of an enemy necessarily a friend, or a friend of a friend a friend? If not, to what extent does this tend to hold? Such questions were formulated in terms of signed (social) networks and necessary and sufficient conditions for a network to be "balanced" were obtained around 1960. Since then the idea that signed networks tend over time to become more balanced has been widely used in several application areas, such as international relations. However investigation of this hypothesis has been complicated by the lack of a standard measure of partial balance, since complete balance is almost never achieved in practice.**

We formalise the concept of a measure of partial balance, compare several known measures on real-world and synthetic datasets, as well as investigating their axiomatic properties. We use both well-known datasets from the sociology literature, such as Read's New Guinean tribes, and much more recent ones involving senate bill co-sponsorship. The synthetic data involves both Erdős-Rényi and Barabási-Albert graphs.

We find that under all our measures, real-world networks are more balanced than random networks. We also show that some measures behave better than others in terms of axioms, computational tractability and ability to differentiate between graphs. We make some recommendations for measures to be used in future work.

I. Introduction

Transitivity of relationships has a pivotal role in analysing social interactions. Is the enemy of enemy a friend? What about friend of an enemy or enemy of a friend? Network science is a key instrument in quantitative analysis for such questions. Researchers of the field are interested in knowing the extent of transitivity of ties and its impact on the global structure and dynamics in communities with positive and negative relationships. Whether the application involves international relationships among states, friendships and enmities between people, or ties of trust and distrust formed among shareholders, relationship to a third entity is always influenced by immediate ties.

There is a growing body of literature that aims to connect theories of social structure with network science tools and techniques to study local behaviours and global structures in signed graphs that come up naturally in many unrelated areas. The building block of structural balance is a work by Heider [1] that was expanded into a set of graph-theoretic concepts by Cartwright and Harary [2] to handle a social psychology problem a decade later. In a setting where the opposite of a negative relationship is a positive relationship, ties to a distant neighbour can be expressed by the product of signs reaching him. The relationship under study has an antonym or dual to be expressed by the opposite sign [3]. Cycles containing an odd number of negative edges are unbalanced, so total balance is guaranteed for signed networks containing no cycles with an odd number of negative edges. This strict condition makes it quite unlikely for a signed network to be totally balanced. The literature on signed networks suggest many different formulas to measure balance. These measures are useful for detecting total balance and total unbalance, but for intermediate cases their performance is not clear and has not been systematically studied.

Our contribution

The main focus of this paper is to provide insight into measuring partial balance, as much uncertainty still exists on this. The dynamics leading to specific global structures in signed networks remain speculative even after studies with fine-grained approaches. The central thesis of this paper is that measures of partial balance should relate to the application, as transitivity is flexible in interpretation in different areas. Thus we need to consider a variety of measures. However, not all measures are equally useful.

We provide a numerical comparison of several measures of partial balance on a variety of undirected signed networks, both randomly generated and inferred from well-known datasets. Using theoretical results for simple classes of graphs, we suggest an axiomatic framework to evaluate these measures and shed light on the context-dependency involved in using such measures.

This paper begins by laying out the theoretical dimensions of the research in Section II and looks at basic definitions and terminology. In Section III different means of checking for total balance are outlined. Section IV discusses some approaches to measure partial balance in Eq. 3 – 11. Section V provides the numerical results in Figures 1 – 6 and Table I. Section VI suggests a set of axioms to evaluate the measures. Section VII is concerned with performance of measures in structures that are close to total balance as in Figure 7. Section VIII presents the findings of the research including Figures 8 – 11, and focuses on context and interpretation. Section IX sums up the research highlights and provides direction for future research. Throughout this paper, the terms signed graph and signed network will be used interchangeably to refer to a graph with positive and negative edges. While several definitions of the concept of balance have been suggested, this paper will only use the original definition of it for undirected signed graphs unless explicitly stated.

II. Problem Statement and Notation

We consider undirected signed networks $G = (V, E, \sigma)$ where σ is the sign function $\sigma : E \to \{-1, +1\}$. The set of nodes is denoted by V, with $|V| = n$. E stands for the set of edges including a total of m edges, m^- negative edges and m^+ positive ones. The expression $u \sim v$ denotes the adjacency of two nodes, regardless of sign. The adjacency matrix is defined in (1). We denote by $|\mathbf{A}|$ the entrywise absolute value of $\mathbf{A}$, which we call the *unsigned adjacency matrix*.

$$\mathbf{A}_{uv} = \begin{cases} \sigma_{u,v} & \text{if } u, v \in E \\ 0 & \text{if } u, v \notin E \end{cases} \tag{1}$$

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, it is a *cycle* of length k. The *weight* of a cycle is the product of the signs of its edges. A cycle is *balanced* if its weight is positive. The total number of balanced cycles (closed-walks) is denoted by O_k^+ (Q_k^+). Respectively, O_k^- (Q_k^-) denotes the total number of unbalanced cycles (closed-walks), and O_k (Q_k) the total number of cycles (closed-walks).

III. Checking for Balance

It is essential to have an algorithmic means of checking for balance. The characterisation of *bi-polarity*, that a signed graph is balanced if and only if its vertex set can be partitioned into two subsets such that each negative edge joins vertices belonging to different subsets, leads to an obvious breadth-first search procedure similar to the usual algorithm for determining whether a graph is bipartite. Moreover, the eigenvalues of signed and unsigned adjacency matrices are equal if and only if the signed network is structurally balanced [4]. Therefore, balance can also be detected by comparing eigenvalues. For our purposes the following additional method is also important. We define the *switching function* g operating over a set of vertices $X \subseteq V$ as follows.

$$\sigma^g_{(u,v)} = \begin{cases} \sigma_{u,v} & \text{if } u, v \in X \text{ or } u, v \notin X \\ -\sigma_{u,v} & \text{if } u \in X \text{ and } v \notin X \text{ or } u \notin X \text{ and } v \in X \end{cases} \tag{2}$$

As the sign of cycles remains the same when g is applied, any balanced graph can switch to an all-positive signature. Accordingly, a balance detection algorithm of complexity $O(n^2)$ can easily be developed by constructing a switching rule on a spanning tree and a root vertex, as suggested in [5].

IV. Measures of Partial Balance

Several ways of measuring the extent to which a graph is balanced have been introduced by researchers. The simplest of such measures is the *degree of balance* suggested by Cartwright and Harary [2], which is simply the fraction of cycles that are balanced:

$$D(G) = \frac{\sum_{k=3}^{n} O_k^+}{\sum_{k=3}^{n} O_k}. \tag{3}$$

Harary [6] provides some properties for the degree of balance in block structures i.e. connected components of network with no cut points. The minimum and maximum values for unbalanced block structures are given by:

$$\frac{m-n}{m-n+2^{m-n}} \leq D(G) \leq \frac{m-n}{m-n+2}. \tag{4}$$

There are two measures closely related to $D(G)$. The first is *relative k-balance* where the sums defining the numerator and denominator of $D(G)$ are restricted to a single term of fixed index k. This extension for $k = 3$ is called the *triangle index*, denoted by $T(G)$.

The second measure can be obtained by weighting cycle lengths as in (5), in which $f(k)$ is a monotonically decreasing function of the length of cycle. The selection of an appropriate function is briefly discussed by Norman and Roberts [7], suggesting functions such as $1/k, 1/k^2, 1/2^k$, but no objective criterion for choosing such a weighting function is known. Exact calculation for cycle-based measures is time-consuming for large networks, as the number of cycles grows exponentially with network size. To tackle the computational complexity, Terzi and Winkler [8] suggested disregarding all cycles longer than three and replacing the remaining triangles by closed-walk of length 3 in calculations as in (6) where $\mathrm{Tr}(A)$ denotes the the trace of matrix. Some researchers argue that dissonance and tension are unclear in cycles of length greater than three [9], justifying the use of the triangle index to analyse structural balance. However the neglected interactions may represent potential tension and dissonance, though not as strong as that represented by unbalanced triads, still determinant of network structure and evolution. Many prefer having a smaller weight for longer cycles, thereby reducing their impact rather than totally disregarding them. Note that $C(G)$ is a generalization of both $D(G)$ and $T(G)$.

$$C(G) = \frac{\sum_{k=3}^{n} f(k) O_k^+}{\sum_{k=3}^{n} f(k) O_k} \tag{5}$$

$$T(G) = \frac{\mathrm{Tr}(A^3) + \mathrm{Tr}(|A|^3)}{2 \times \mathrm{Tr}(|A|^3)} \tag{6}$$

Beside checking cycles, there are computationally easier approaches to structural balance such as the walk-based approach. A walk-based measure of balance is suggested by Estrada and Benzi [10] with more weight placed on shorter closed-walks than the longer ones. They have tested their measure on five signed networks resulting in values inclined towards unbalance. Moreover, they argued that balance measures are sensitive to the weight of social links, resulting in the conflicting observations prior to their study [10]. According to their measure, comparison of higher powers of the signed and unsigned adjacency matrices reveals some insight into network structural properties.

Let $\mathrm{Tr}(e^A)$ and $\mathrm{Tr}(e^{|A|})$ denote the trace of matrix exponential for $\mathbf{A}$ and $|\mathbf{A}|$ respectively. In this formula, closed-walks are weighted by the inverse factorial of their lengths which has a relatively fast rate of decay comparing to weighting functions previously suggested for cycle-based measures. The weighted ratio of balanced to total closed-walks is formulated as follows:

$$W(G) = \frac{2K}{K+1}, K = \frac{\sum_k \frac{Q_k^+ - Q_k^-}{k!}}{\sum_k \frac{Q_k^+ + Q_k^-}{k!}} = \frac{\mathrm{Tr}(e^A)}{\mathrm{Tr}(e^{|A|})} \tag{7}$$

A clustering-based measure for balance is suggested by Kunegis [11] taking insight from the classic clustering coefficient denoted by $CC(G)$. The signed clustering coefficient, $SC(G)$, is defined similarly, but it numerically approximates the number of times that the closing edge creates a balanced triad as in (8).

$$CC(G) = \frac{\{u, v, w \in V | u \sim v \sim w \sim u\}}{\{u, v, w \in V | u \sim v \sim w\}}$$

$$SC(G) = \frac{\sum_{u \sim v \sim w \sim u} \sigma(\{u,v\})\sigma(\{v,w\})\sigma(\{w,u\})}{\{u, v, w \in V | u \sim v \sim w\}} \tag{8}$$

Additionally, the normalised relative signed clustering coefficient is defined as in (9). This measure is an approximation of the triangle index based on sampling over a fixed number of triads in the graph. A drawback of this measure is that it disregards longer cycles like [8]. On the other hand, it can be computed readily for large networks. As balanced (unbalanced) sampled triads contribute to $S(G)$ positively (negatively), it ranges between $[0,1]$ and is comparable to the other measures.

$$S(G) = \frac{SC(G) + CC(G)}{2 \times CC(G)} \approx \frac{O_3^+}{O_3} = T(G) \tag{9}$$

A quite different measure is the *frustration index*. Originally proposed for applications on ferromagnetic molecules, it is also referred to as the *line index for balance* by [12]. It equals the minimum number of edges whose deletion (or equivalently, negation) result in a balanced graph. In a setting where each vertex is given a value of ± 1 as well, if the endpoints of positive (negative) edges are in the same (opposite) states, they are satisfied, while edges violating such rules are “frustrated". The frustration index is therefore the smallest number of frustrated edges over all possible assignments of ± 1 to the nodes. Similarly, this measure equals the number of members in a collection of edges, called *deletion-minimal*, whose deletion results in balance, while there is no subset of this collection yielding balance. Each edge in a deletion-minimal set lies on an unbalanced cycle and every unbalanced cycle of the network contains an odd number of edges of the deletion-minimal set. This measure is hard to compute as the problem can be reduced to graph maximum cut problem, in a special case of all negative edges, which is known to be NP-hard. However, upper bounds can be readily provided for line index for balance, as denoted by $L(G)$, such as $L(G) \leq m^-$ which states the obvious result of removing all negative edges.

The number of frustrated edges in Erdős-Rényi graphs with equal probabilities for positive and negative signs are analysed by El Maftouhi, Manoussakis and Megalakaki [13]. It follows a binomial distribution with parameters $n(n-1)/2$ and $p/2$. Therefore, the expected value for frustrated edges is $n(n-1)p/4$. They also prove that such a network is almost always not balanced when $p \geq \log 2/n$.

It is straightforward to prove that frustration index is equal to the minimum number of negative edges over all switching functions [5]. Moreover, if $m^-(G^g) = L(G)$ then every vertex under this switching satisfies $d^-(v^g) \leq d^+(v^g)$. Tomescu [14] proves that this measure is bounded by $\lfloor (n-1)^2/4 \rfloor$. Bounds for the largest frustration index over all signings of vertices are provided by [15]:

$$\frac{m}{2} - \sqrt{mn} \leq \max L(G) \leq \frac{m}{2}. \tag{10}$$

An exhaustive search algorithm can be used for flipping edges and calculating frustration index. In order to compare with the other indices which take values in the unit interval and give the value 1 for balanced graphs, we normalise the frustration index by dividing by the maximum possible value and map it via a decreasing function. This yields the *normalised frustration index*, which we denote by $F(G)$:

$$F(G) = 1 - \frac{2L(G)}{m}. \tag{11}$$

Balance in signed networks can be seen through another view as well. Vertices can be grouped into increasingly homogeneous positions by iteratively calculating Pearson moment correlations from the columns of the adjacency matrix [16]. This reveals within-block and between-block connections in the reduced-form matrix. Assuming position as a set of vertices, blocks are sets of ties between positions. Generalized blockmodeling reveals network structural properties such as balance. Although perfect balance is unlikely, networks may have partitions that are close to perfectly balanced [16]. Doreian and Mrvar [17] discuss this approach in partitioning signed social networks extensively.

V. Numerical Results

Measures of partial balance, denoted by $\mu(G)$ are calculated for both Erdős-Rényi and Barabási-Albert random networks. The same randomly generated graphs with different number of negative edges assigned on random are used to analyse balance. Figures 1 – 2 demonstrate the partial balance in random networks measured by different methods. As it is shown measures have different sensitivity to the number of negative edges. Interestingly, the degree of balance, $D(G)$, is observed to tend to 0.5 for random networks with $m^- \geq 7$. No difference to $D(G)$ is observed where $C(G)$ is weighted by $f(k) = 1/k$ as values are almost equal. Neither one of them differentiates partial balance in networks with non-trivial number of negative edges $m^- \geq 7$. From the charts, it can be seen that $C(G)$ weighted by $f(k) = 1/k!$ decreases slower than the former two and then fluctuates around 0.5 for $m^- \geq 15$. Moreover, a steady linear decrease is observed from $F(G)$ for $m^- \leq 10$ and then it fluctuates around 0.5 for networks with greater number of negative edges. The single most striking observation to emerge is from the walk-based measure, $W(G)$, that drops to 0.3 for $m^- \geq 10$ and then decreases to 0.1 for networks with more negative edges. Triangle index, $T(G)$, and its approximation, $S(G)$, are the two measures with the widest range of values almost steadily decreasing to 0.5 where $m^- \leq 15$ then fluctuating around it for $15 \leq m^- \leq 35$ and finally decreasing to the smallest values on the charts where $m^- \geq 35$.

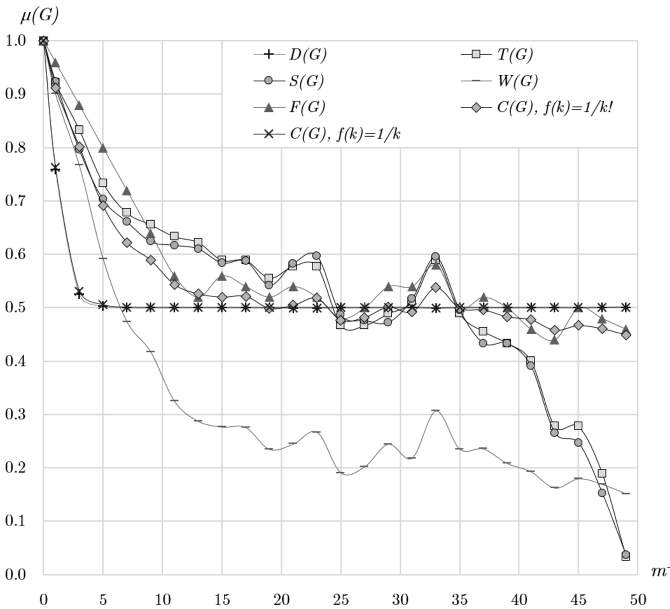

Fig. 1. Partial balance measured by different methods for Erdős-Rényi (n=15, p=0.45) network with 50 edges and 2573532 cycles

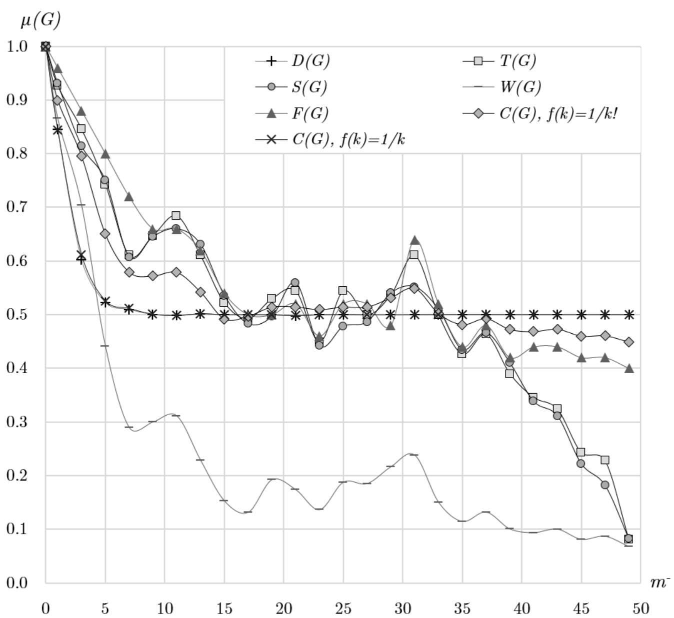

Fig. 2. Partial balance measured by different methods for Barabási-Albert Preferential Attachment (n=15,m=5) network with 50 edges and 411890 cycles

In the remaining part of this section, we report measures of partial balance for a range of small signed networks inferred from datasets. There are well-studied datasets on small communities with positive and negative interactions and preferences. Read's dataset for New Guinean highland tribes [18] is demonstrated as a signed graph in Figure 3, where dotted lines represent negative edges and solid lines represent positive ones. Sampson's dataset for monastery interactions [19] is drawn in Figure (4). There are also datasets of students' choice and rejection [20], [21] as demonstrated in Figure 5 and Figure 6. The last three are converted to signed graphs by considering mutually agreed relations. Moreover, a larger signed network is inferred by [22] through implementing a stochastic degree sequence model on Fowler's data on Senate bill co-sponsorship [23]. The results are shown in Table I where measures for the random graphs with the same parameters are also provided for comparison.

The cycle based measures are difficult to compute in large networks with more than 10^9 cycles. The degree of balance provides useless values clustered around 0.5 that is not unusual. Its extended measure weighted by $f(k) = 1/k$ provides the same values that are not reported in the table. However, $C(G)$ with $f(k) = 1/k!$ provides much more relevant values making real networks and random graphs distinguishable. $T(G)$, $S(G)$, and $F(G)$ also give reasonable values to distinguish partial balance in real networks and their reciprocal random graphs. Although neither of the networks are completely balanced, small values of $L(G)$ suggests that removal of only a few edges makes the real networks completely balanced. From this data, we can see that random networks have lower partial balance. It indicates that ties in real signed networks are more transitive than what we expect by random. A clear reason for $W(G)$ providing small values for senate network could not be identified in this analysis. This raises questions about potential dependency of $W(G)$ to network parameters which will be discussed in the next Section.

VI. Axiomatic framework of evaluation

Two different sets of axioms are suggested in [7], which characterise the measure $C(G)$ (up to the choice of $f(k)$). Moreover, the theory of structural balance itself is axomatised in [24]. Here we provide another set of axioms for measures of partial balance in order to shed light on their applications and context. We define a *measure of partial balance* to be a function μ taking each signed graph to an element of $[0,1]$. However, worthy of mention is that some of these measures were originally defined as a measure of unbalance (frustration index and the original walk-based measure suggested by [10]) calibrated at 0 for completely balanced structures, so that some normalisation was required, and perhaps our normalisation choices can be improved on.

We list the following axioms.

- A1 $\mu(G) = 1$ if and only if G is balanced.
- A2 $\mu(G) = \mu(G \oplus G)$, where the graph on the right denotes the disjoint union of two copies of G.
- A3 If C is a cycle with positive weight, then $\mu(G) \leq \mu(G \oplus C)$.
- A4 If C is a cycle with negative weight, then $\mu(G) \geq \mu(G \oplus C)$.

The first three axioms holds in all the measures introduced. Following the addition of a negative cycle, $W(G)$ is observed to increase resulting in its failure in the last axiom (for example, take $G = K_5$ with single negative edge, and C to be of length 3 cycle with a single negative edge). Moreover, $F(G)$ satisfies the fourth axiom whenever $L(G) \leq m/k$ where k is the length of C, but it is easy to see by considering

TABLE I. MEASURES OF PARTIAL BALANCE CALCULATED FOR FIVE SIGNED NETWORKS AND THEIR SIMILAR RANDOM GRAPHS

Graph	n	m	m^-	Cycles	D(G)	C(G)	W(G)	T(G)	S(G)	F(G)	L(G)
Highland tribes [18]	16	58	29	22216973	0.491	0.677	0.527	0.868	0.884	0.759	7
G(n,m)				31791520	0.500	0.496	0.173	0.516	0.521	0.448	16
Monastery Interactions [19]	18	49	12	1436972	0.504	0.717	0.743	0.857	0.841	0.796	5
G(n,m)				1347065	0.500	0.536	0.497	0.567	0.549	0.633	9
Fraternity preference [20]	17	40	17	107928	0.498	0.739	0.835	0.778	0.801	0.800	4
G(n,m)				110152	0.498	0.511	0.685	0.500	0.487	0.650	7
College preference [21]	17	36	16	15265	0.498	0.616	0.764	0.786	0.841	0.667	6
G(n,m)				19289	0.500	0.442	0.601	0.417	0.541	0.611	7
Senate bill co-sponsorship [22]	100	2461	1047	$\geq 10^9$			0.002	0.864	0.843	0.731	331
G(n,m)				$\geq 10^9$			0.000	0.507	0.522	≥ 0.216	≤ 965

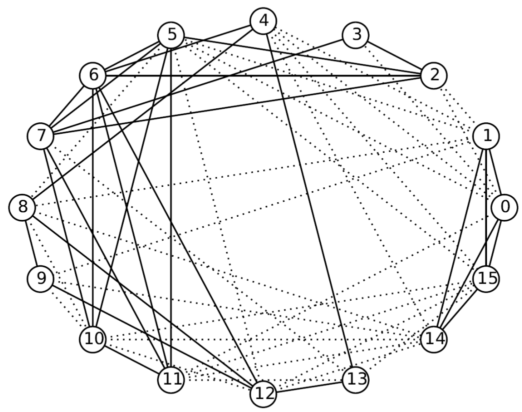

Fig. 3. Highland tribes network, a signed network of 16 tribes of the Eastern Central Highlands of New Guinea [18]

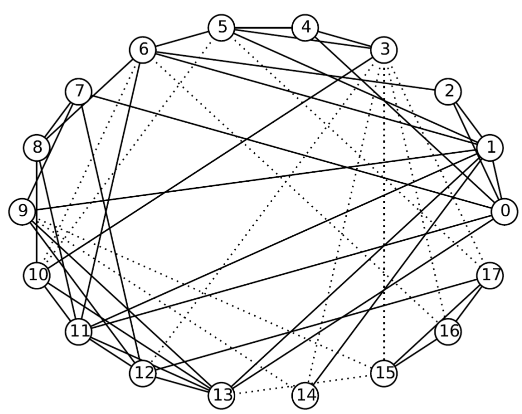

Fig. 4. Monastery interactions network of 18 New England novitiates inferred from the integration of all positive and negative relationships in time T4 [19]

a sufficiently long negative cycle that $F(G)$ fails the fourth axiom in general.

Another desirable property which we have not formulated as a formal requirement owing to its vagueness, is that the measure takes on a wide range of values. For example, $W(G)$ tends to 0 as $|V(G)|$ increases, for some families of graphs such as the Senate database in Table I, which makes comparison with other measures difficult. It also has a vastly different behaviour for another family of graphs that will be discussed in the next section.

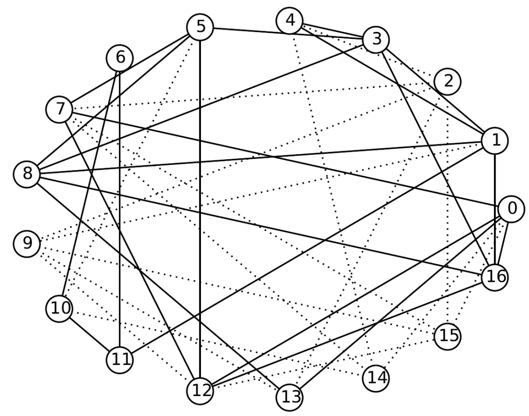

Fig. 5. Fraternity preferences network of 17 boys living in a pseudo-dormitory inferred from ranking data of the last week in [20]

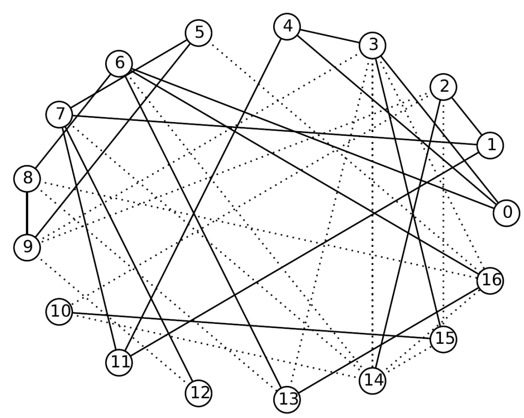

Fig. 6. College preferences network of 17 girls at an Eastern college inferred from ranking data of house B in [21]

VII. ALMOST BALANCED NETWORKS

In this section we analyse the capability of measures in measuring partial balance in structures that are only one edge away from a state of total balance. Particularly, we are interested in complete graphs K_n with one negative edge. It is straight-forward to provide closed-form formula for different measures of partial balance in such nicely structured graphs. Measures of partial balance, $D(K_n), C(K_n), T(K_n), W(K_n)$ and $F(K_n)$ can be expressed as in (12) – (16) (details of calculations are provided in the appendix).

$$D(K_n) = 1 - \frac{2}{n(n-1)} \frac{\sum_{k=3}^{n} \frac{n!}{(n-k)!}}{\sum_{k=3}^{n} \frac{n!}{(n-k)!k}} \tag{12}$$

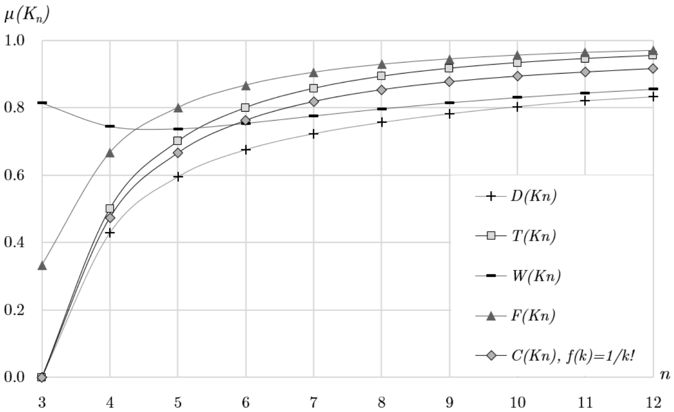

Fig. 7. Partial balance measured by different methods for K_n with one negative edge i.e. complete graphs where $L(K_n) = 1$

$$C(K_n) = 1 - \frac{2}{n(n-1)} \frac{\sum_{k=3}^{n} \frac{n!}{k!(n-k)!}}{\sum_{k=3}^{n} \frac{n!}{k!(n-k)!k}} \tag{13}$$

$$T(K_n) = 1 - \frac{6}{n(n-1)} \tag{14}$$

$$W(K_n) \sim \frac{2e^{-4/n}}{1+e^{-4/n}} \tag{15}$$

$$F(K_n) = 1 - \frac{4}{n(n-1)} \tag{16}$$

Figure 7 demonstrates the behaviour of different indices for K_n with a single negative edge. $W(K_n)$ gives unreasonably large values for $n \leq 5$. $D(K_n)$ can be observed to tend to a value significantly smaller than 1 for such networks with large n. Values of $C(K_n)$ are almost always larger than $D(K_n)$. $T(K_n)$ provides the most relevant values while $F(K_n)$ gives larger values and can be observed to measure partial balance differently as it is not based on cycles.

VIII. Discussion

Figures 8 – 11 show the four small signed networks, with a minimal deletion set of edges indicated by dotdash lines. It is interesting to see how these small signed networks can be made balanced by removing a few edges. It shows how such networks are not that far from balance (when measured by normalised frustration index), contrary to some of the measures we calculated.

One criticism of much of the literature on balance theory is that it is widely used on directed signed graphs. It seems that this approach is questionable in two ways. First, it neglects the fact that many edges in signed digraphs are not reciprocated. Therefore, one side of such relationships is not aware of good will or ill will towards him. Implementing balance theory on signed digraphs provides the extend of dissonance avoidance no matter known or unknown. Secondly, it does not make use of the directionality of ties and the concepts of sending and receiving positive and negative links. Moreover, some issues stay unclear when extending balance theory to digraphs by replacing cycles with semicycles. For instance, the signed digraph in question may have two arcs with opposite signs between two nodes. The two arcs are simply replaced with an edge, but the signs of edge is undefined.

In a parallel line of research on network structural analysis, researchers differentiate between classical balance theory and structural balance specifically in the way that the latter is directional [9]. They consider another setting for defining balance where absence of ties implies negative relationships. This assumption makes the theory limited to complete signed digraphs. Accordingly, 64 possible structural configurations emerge for three nodes. These configurations can be reduced to 16 classes of triads, referred to as 16 MAN triad census, based on the number of Mutual, Asymmetric, and Null relationships they contain. There are only 2 out of 16 classes that are considered balanced. New definitions are suggested by researchers in order to make balance theory work in a directional context. According to Prell [25], there is a second, a third, and a fourth definition of permissible triads allowing for 3, 7, and 9 classes of all 16 MAN triads. However, there have been many instances of findings in conflict with expectations [25].

Leskovec, Huttunlocher and Kleinberg [26] compare the reliability of predictions made by competing theories of social structure: balance theory and status theory (a theory that explicitly includes direction and gives quite different predictions). The consistency of these theories with observations is investigated through large signed directed networks such as Epinions, Slashdot, and Wikipedia. The results suggest that status theory predicts accurately most of the time while predictions made by balance theory are incorrect half of the time. This supports the inappropriateness of balance theory for structural analysis of signed digraphs.

Apart from directionality, the interpretation of balance measures is highly essential. Numerous studies have compared balance measures with their extremal values and found that signed networks are far from balanced. However, with such a strict criterion, caution must be applied not to look for properties that are almost impossible. A much more systematic approach would identify how measures of balance interact with other network parameters that are linked to transitivity of ties, such as number of nodes, positive edges, and negative edges. According to this approach, we compare with the value of the balance measures for an analogous random graph with the same network parameters. Table I provides this comparison showing that the extent of transitivity of ties in signed networks is substantial. The real signed networks analysed are more balanced than we expect by chance. Therefore random signed networks with non-trivial number of negative edges appear to be less balanced than real-world signed networks, while there are extremal cases of total balance on one side and deliberately designed unbalanced structures on the other side.

Taken together, the findings do not support strong recommendations to use cycle-based measures, as they are difficult to compute and do not provide a proper range of values, whether weighted or not. The triangle index, however, seems to behave well, and is simple to compute because closed walks of length 3 are the same as (directed) 3-cycles. Moreover, networks can be differentiated by the wide range of values that $T(G)$ provides. $S(G)$ can be used

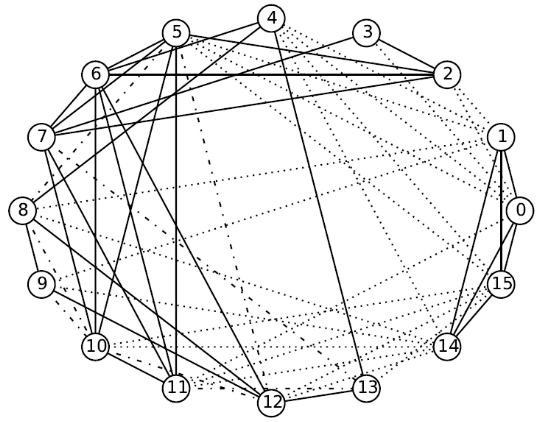

Fig. 8. Highland tribes network becomes balanced after removing 7 negative edges

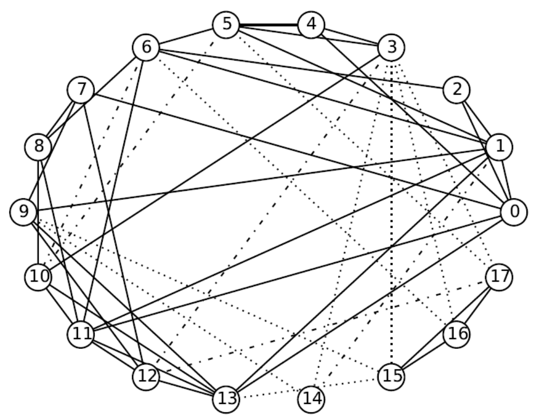

Fig. 9. Monastery interactions network becomes balanced after removing 2 positive and 3 negative edges

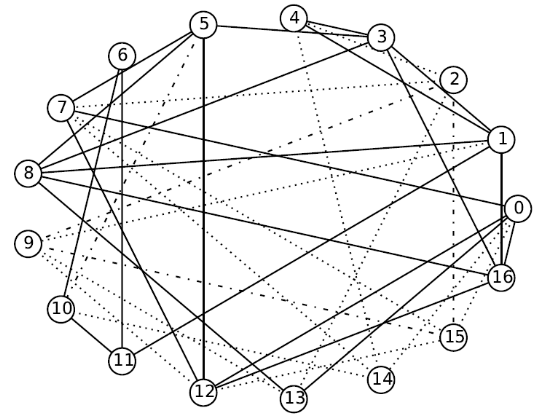

Fig. 10. Fraternity preferences network becomes balanced after removing 4 negative edges

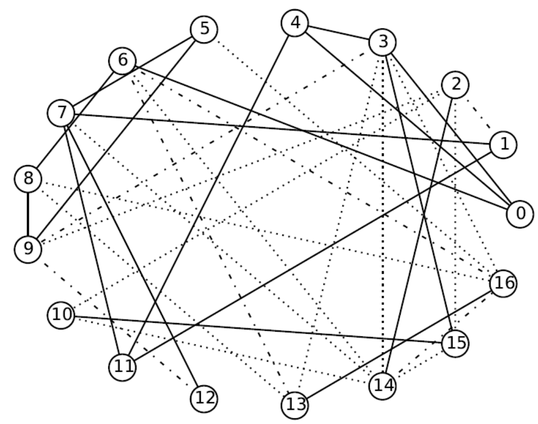

Fig. 11. College preferences network becomes balanced after removing 3 positive and 3 negative edges

as its approximation in case of large networks. Continued efforts are needed to make frustration-based measures comparable to the others while satisfying all four axioms introduced. Frustration seems to measure something different from cycle balance, and be worth pursuing in future. Walk-based measures like $W(G)$ can perhaps be improved by a more systematic way of weighting closed-walks to avoid double-counting of cycles and closed-walks with repeated edges. However the clustering of values near zero for large networks may present problems.

IX. Conclusion and Future Research

Returning to the questions posed at the beginning of this study, it is now possible to state that many signed networks exhibit a level of partial (but not total) balance beyond that expected by chance. One of the more significant findings to emerge from this study is that methods suggested for measuring balance have their context and interpretation. Although this study focuses on partial balance, the findings may well have a bearing on link prediction and clustering in signed networks. The present study confirms previous findings that theory of structural balance cannot be taken as a reliable predictor of network evolution. It contributes additional evidence that suggests a gray-scale for transitivity of positive and negative relationships. Although the major part of the current study is based on small signed networks, the findings suggest the inefficacy of some methods for analysing larger networks as well. One gap in this study which could have affected the measurements of partial balance is that we avoid using structural balance theory for analysing directed networks, making a significant part of the literature untested by our approach for now. However, see our discussion in Section VIII.

The findings of this study have a number of important implications for future investigation. Having an efficient measure of partial balance in hand, we plan to analyse international relations. The literature in international relations offers data on formal alliances between some countries as well as hostile relations between some others. Middle East is an interesting region due to the number of conflicts among states of a region. A signed graph can be readily designed where balance theory and frustration index predict which relations need to be protected to avoid a cold war bi-polarity situation in the region.

Acknowledgment

The authors thank Te Pūnaha Matatini for their support.

APPENDIX: DETAILS OF CALCULATIONS

In K_n with one negative edge, a k-cycle is specified by choosing k vertices in some order, then correcting for the overcounting by dividing by 2 (the possible directions) and k (the number of starting points, namely the length of the cycle). If the unique negative edge is required to belong to the cycle, we need choose only $k-2$ further elements and no overcounting occurs. Therefore, the number of negative cycles and total cycles are given by:

$$\sum_{k=3}^{n} O_k^- = \sum_{k=3}^{n} \frac{(n-2)!}{(n-k)!}, \quad \sum_{k=3}^{n} O_k = \sum_{k=3}^{n} \frac{n!}{2k(n-k)!}.$$

Similarly, $T(K_n)$ can be calculated as

$$T(K_n) = 1 - \frac{\frac{(n-2)!}{(n-3)!}}{\frac{n!}{2\times 3(n-3)!}} = 1 - \frac{6}{n(n-1)}.$$

As in such a graph $L(K_n) = 1$, the normalised frustration index can be expressed as

$$F(K_n) = 1 - \frac{2L(K_n)}{m} = 1 - \frac{2}{n(n-1)/2} = 1 - \frac{4}{n(n-1)}.$$

The unsigned adjacency matrix $|\mathbf{A}|$ of the complete graph has the form $E - I$ where E is the matrix of all 1's. The latter matrix has rank 1 and nonzero eigenvalue n. Thus $|\mathbf{A}|$ has eigenvalues $n-1$ (with multiplicity 1) and -1 (with multiplicity $n-1$). The matrix $\mathbf{A}$ has a similar form and we can guess eigenvectors of the form $(-1,1,0,\ldots,0)$ and $(a,a,1,1,\ldots,1)$. Then a satisfies a quadratic $2a^2+(n-3)a-(n-2)=0$. Solving for a and the corresponding eigenvalues, we obtain eigenvalues $(n-4\pm\sqrt{n-2)(n+6)})/2, 1, -1)$ (with multiplicity $n-3$). This yields

$$K = \frac{(n-3)e^{-1} + e + e^{\frac{n-4-\sqrt{(n-2)(n+6)}}{2}} + e^{\frac{n-4+\sqrt{(n-2)(n+6)}}{2}}}{(n-1)e^{-1} + e^{n-1}}$$

which results in $W(K_n) \sim \frac{2e^{-4/n}}{1+e^{-4/n}}$.

REFERENCES

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

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

[3] F. Harary, "Structural duality," *Behavioral Science*, vol. 2, no. 4, pp. 255–265, 1957. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/bs.3830020403/abstract

[4] B. D. Acharya, "Spectral criterion for cycle balance in networks," *Journal of Graph Theory*, vol. 4, no. 1, pp. 1–11, 1980. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/jgt.3190040102/abstract

[5] T. Zaslavsky and C. R. Rao, "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

[6] 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

[7] 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

[8] 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

[9] P. Bonacich, *Introduction to mathematical sociology.* Princeton: Princeton University Press c2012, 2012.

[10] 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, p. 042802, 2014. [Online]. Available: http://journals.aps.org/pre/abstract/10.1103/PhysRevE.90.042802

[11] 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

[12] F. Harary, "A matrix criterion for structural balance," *Naval Research Logistics Quarterly*, vol. 7, no. 2, pp. 195–199, Jun. 1960. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/nav.3800070208/abstract

[13] A. El Maftouhi, Y. Manoussakis, and O. Megalakaki, "Balance in random signed graphs," *Internet Mathematics*, vol. 8, no. 4, pp. 364–380, 2012. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/15427951.2012.675413

[14] 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

[15] 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

[16] P. Doreian, *Generalized blockmodeling*, ser. Structural analysis in the social sciences ; 25. Cambridge ; New York: Cambridge University Press 2005, 2005.

[17] P. Doreian and A. Mrvar, "Partitioning signed social networks," *Social Networks*, vol. 31, no. 1, pp. 1–11, Jan. 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378873308000397

[18] 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

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

[20] 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.

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

[22] 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

[23] 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

[24] T. Schwartz, "The friend of my enemy is my enemy, the enemy of my enemy is my friend: Axioms for structural balance and bi-polarity," *Mathematical Social Sciences*, vol. 60, no. 1, pp. 39–45, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016548961000020X

[25] C. Prell, *Social network analysis : history, theory and methodology.* London ; Los Angeles: SAGE 2012, 2012.

[26] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 2010, pp. 1361–1370.