3

Measuring Balance of Signed Networks and its Application to Sign Prediction

In this chapter, we will use the term "network" for "graph" and by balance we mean strong balance. We first show that the walk based metric introduced in [Estrada and Benzi, 2014] for the detection of "lack of balance" in social networks could be quite misleading as it tells that large real networks, for example, the empirical networks Wikielection, Slashdot, and Epinion are 100% unbalanced. We justify our arguments by showing that this happens due to the curse of the formulation of the metric proposed in [Estrada and Benzi, 2014]. Thus we introduce a new parameterized metric for the measure of the degree of balance of signed networks by using weighted closed walks and the Katz measure of similarity. This, in contrast to the claim in [Estrada and Benzi, 2014], shows that the real world signed networks are not 100% unbalanced in fact it all depends upon what weights are being used for the closed walks of finite length in a walk-based metric. The proposed measure in this chapter also contradicts the claim made in [Estrada and Benzi, 2014] that large signed random networks are 100 % unbalanced.

Here we mention that while there is no method to be sure whether a new measure better quantifies the balance of the empirical networks, since the absolute balance is not universally defined and is not known, nevertheless, one can compare the performance of the old and new measures on randomly generated networks. For instance, if the old measure cannot distinguish two networks having approximately same number of vertices and edges generated by using different values of q (the probability for negative edges), while the new measure gives different values of lack of balance for these two networks, then it objectively shows that the new measure provides a better quantification of balance. Besides, in that case, the new measure determines structural dissimilarity of two random networks of approximately same size.

We employ the proposed metric to three different random signed networks having an approximately same number of vertices and edges generated by considering different probabilities for the creation of negative edges and show that their lack of balance differs. Whereas, the measure proposed in [Estrada and Benzi, 2014] can not distinguish the lack of balance in these networks and show that they are 100% unbalanced. Finally, we consider the problem of sign prediction in signed networks that deal with predicting the sign of an edge by using the signs of edges in the rest of network. Here, we use the well-known Katz prediction rule as discussed in [Chiang et al., 2011]. Thus we conclude that not the longer cycles but the use of cycles of lengths $4, 5, 6$ can better predict the sign of edges in the signed networks considered in this chapter.

3.1 WALK-BASED MEASURE FOR DEGREE OF LACK OF BALANCE

Walk-based measures to study structural properties of networks have become popular after the success of the idea of communicability for unsigned networks [Estrada and Hatano, 2008] [Crofts and Higham, 2009]. It would not be exaggerated to say that the idea of communicability has been exploited to introduce the walk-based measure for lack of balance for signed networks in [Estrada and Benzi, 2014]. In this section, we briefly review the walk-based measure for balance which was introduced in [Estrada and Benzi, 2014] and provide a mathematical reasoning how it led to conclude that real networks are poorly balanced. First, we recall the following preliminaries.

Let $G = (V, E)$ be a signed graph. The adjacency matrix $A = (a_{ij})$ of order $|V| \times |V|$ associated with G is given by

$$a_{ij} = \begin{cases} 1 & \text{if } (i,j) \in E_+ \\ -1 & \text{if } (i,j) \in E_- \\ 0 & \text{if } (i,j) \notin E \end{cases}$$

where E_+ and E_- denote the set of positive and negative edges of G respectively, such that $E = E_+ \cup E_-$. A walk of length k in G is a sequence of (not necessarily distinct) vertices $v_1, v_2, \ldots, v_{k-1}, v_k$ such that for each $i = 1, 2, \ldots, k-1$, there is a edge from v_i to v_{i+1}. If all vertices are distinct in a walk then the walk is called a path. If $v_k = v_1$ the walk (path) is called a closed walk (cycle). In addition, the sign of a walk is defined as the product of the signs of its edges. Similar to triads, a walk is called balanced if its sign is positive, otherwise it is called unbalanced.

Further, every signed network has an underlying unsigned network which consists of the same set of vertices and edges as G and all edges are having a positive sign. Let us represent the underlying network of G by $|G|$. Adjacency matrices of G and $|G|$ are denoted by A and $|A|$ respectively. Evidently, the total number of closed walks of length k in G is given by trace(A^k). A Balanced Weighted Closed Walk (BCW) is a positively signed closed walk of nonzero length. Similarly, an Unbalanced Weighted Closed Walk (UCW) is a negatively signed closed walk of nonzero length.

The walk-based measure for the degree of lack of balance in a signed network G on n vertices is defined by

$$U = \frac{1-K}{1+K}, \text{ where } K = \frac{\text{trace}(e^A)}{\text{trace}(e^{|A|})} = \frac{\sum_{j=1}^n \exp(\lambda_j(G))}{\sum_{j=1}^n \exp(\lambda_j(|G|))}, \tag{3.1}$$

$\lambda_j(G)$ and $\lambda_j(|G|)$, $j = 1, \ldots, n$ are eigenvalues of A and $|A|$ respectively, in ascending order [Estrada and Benzi, 2014]. Thus U can be interpreted as the ratio of UCWs to BCWs. Note that, in calculating the lack of balance the weights of an m length walk is assumed to be $1/m!$ which is a decreasing function of length. Note also that this measure (3.1) has a resemblance of the measure (2.2). It follows from the Theorem 2.2, when $\{\lambda_j(G) : j = 1, \ldots, n\} = \{\lambda_j(|G|) : j = 1, \ldots, n\}$, the network is balanced and $U = 0$. On the other hand, when the graph is highly unbalanced, $U \approx 1$, that is $\sum_{j=1}^n \exp(\lambda_j(G)) \ll \sum_{j=1}^n \exp(\lambda_j(|G|))$.

3.1.1 Limitations of this method for undirected large signed networks

It is not obvious from the definition of K why the weight $\frac{1}{m!}$ is considered for a closed walk of length m. Nonetheless, it gives a compact representation of U. When this measure is applied to the real world networks Slashdot, Epinions, and WikiElection, it shows these are 100% unbalaced networks [Table II, [Estrada and Benzi, 2014]]. This result has been justified by supposing "the triads with only one negative edge or with all three negative edges have been found to be overrepresented in the three online networks" which is an observation made in [Leskovec et al., 2010]. However, it follows from the definition of K that

$$K = \frac{\sum_{j=1}^n \exp(\lambda_j(G))}{\sum_{j=1}^n \exp(\lambda_j(|G|))} = \frac{\exp(\lambda_n(G))}{\exp(\lambda_n(|G|))} \frac{\mathcal{O}(n)}{\mathcal{O}(n)} \tag{3.2}$$

and from the Table 3.1, $\exp(\lambda_n(G)) \ll \exp(\lambda_n(|G|))$ which finally conclude that $K \approx 0$. Let $\rho(G)$ denote the spectral radius of G, that is, the maximum absolute value of the eigenvalues of G. We

Table 3.1 : n: number of vertices, m^+: number of positive edges, m^-: number of negative edges, $\triangle^+$: number of balanced triangles, $\triangle^-$: number of unbalanced triangles, κ: edge density $=\frac{2(m^+ + m^-)}{n(n-1)}$, $||A||_\infty$ is infinity norm of adjacency matrix of G, RN: random network.

Networks	n	m^+	m^-	$\triangle^+$	$\triangle^-$	κ	$\lambda_n(\|G\|)$	$\lambda_n(G)$
$Wikielection$	7118	92238	7784	651560	72398	0.0039	142.7757	130.6673
$Slashdot$	9000	75462	25707	226044	31228	0.0025	104.53	95.9208
$Epinions$	8000	91498	14308	713563	147369	0.0033	164.0028	138.7156
$RN-I$	8000	319760	15979	171625	25528	0.0105	84.94	77.06
$RN-II$	8000	320426	31872	88072	25644	0.0110	89.0591	73.3225
$RN-III$	8000	319180	159953	148767	137909	0.0150	120.7704	42.8063
WWI	n	m^+	m^-	$\triangle^+$	$\triangle^-$	$\|\|A\|\|_\infty$	$\lambda_n(\|G\|)$	$\lambda_n(G)$
$ThreeEmperorsLeague$	6	3	6	5	2	4	3.6458	3.1028
$TripleAlliance$	6	5	6	6	2	5	3.8590	3.4163
$German-RussianLapse$	6	3	7	3	2	5	3.5141	3.0144
$French-RussianAlliance$	6	4	7	6	2	5	3.8590	3.3743
$EntenteCordiale$	6	5	6	6	2	5	3.8590	3.4163
$British-RussianAlliance$	6	6	9	20	0	5	5	5

mention that all the large signed networks including the random networks that are considered in this chapter and their underlying positive networks have spectral radii equal to the corresponding largest eigenvalue. In general, for any signed network G for which the spectral radius $\rho(G) = \lambda_n(G)$, $K \approx 0$ follows by the fact that $\rho(G) < \rho(|G|)$ [page 619, [Meyer, 2000]] when $\rho(G) \neq \rho(|G|)$. In addition, if $\lambda_n(G) = \rho(G) = \rho(|G|) = \lambda_n(|G|)$, the balance of G increases depending on the distribution of eigenvalues of G and $|G|$, by equation (3.2).

Since, the connection between balance of a signed network and the spectral radius of the network is not known, in fact it is difficult to find such a connection, we conclude that the 100% unbalance of the real world networks Slashdot, Epinions, and WikiElection is due to the curse of this measure not due to the structural properties of these networks. This calls for the development of new potential measures for quantifying lack of balance in signed networks.

3.2 PARAMETERIZED WALK-BASED MEASURE FOR LACK OF BALANCE

In this subsection, we propose a parameterized walk-based measure by exploiting the concept of Katz index popularly used for finding the similarity of vertices in unsigned networks. Katz measure is also valid for signed networks and gainfully used for edge prediction problem in signed networks [Chiang et al., 2014]. Note that, this is a resolvent centrality measure defined as

$$\sum_{l=0}^{\infty} \beta^l A_{ij}^l = \left((I - \beta A)^{-1}\right)_{ij}, \tag{3.3}$$

where A_{ij}^l is the number of closed walks of length l between the vertices i and j and $\beta \in [0,\ 1/\rho(A)]$, where $\rho(A)$ denotes the spectral radius of A which is the adjacency matrix corresponding to a signed network G. It is needless to mention that A_{ii}^l provides the number of closed walks of length l adjacent to the vertex i and $\left((I-\beta A)^{-1}\right)_{ii}$ provides the weighted sum of the number of closed walks at the vertex i, with walks of length l scaled by β^l. We mention that this centrality measure has been considered as a measure of imbalance for signed networks in [Chiang et al., 2011].

Inspecting the definition given in (3.1) it will be tempting to define a measure for the degree of balance of a signed network by using Katz measure as

$$K_z = \frac{\sum_{i=1}^{n}\sum_{l=0}^{\infty}\beta^l A_{ii}^l}{\sum_{i=1}^{n}\sum_{l=0}^{\infty}\beta^l |A|_{ii}^l}, \beta \in [0,\ 1/\rho(|A|)]$$

which is a valid mathematical definition as $\rho(A) \le \rho(|A|)$ and n is the number of vertices in the network. However, observe that, whatever large the network is, the same edges will get counted a large number of times for calculating closed walks described by A_{ii}^l as the value of l increases. On the other hand, since we are only interested in closed cycles in the network, it is of no use to include the terms corresponding to $l = 0, 1, 2$ in the definition of K_z. Hence, we define a measure for the degree of structural balance of a signed network as

$$\begin{aligned} K(\beta,k) &= \frac{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^l A_{ii}^l}{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^l |A|_{ii}^l} = \frac{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^l \lambda_i(G)^l}{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^l \lambda_i(|G|)^l} \\ &= \frac{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^{l-3} \lambda_i(G)^l}{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^{l-3} \lambda_i(|G|)^l} = \frac{\sum_{l=3}^{k}\beta^{l-3}\mathrm{Tr}(A^l)}{\sum_{l=3}^{k}\beta^{l-3}\mathrm{Tr}(|A|^l)}, \end{aligned} \tag{3.4}$$

where $\beta > 0$, Tr denotes trace, and k is the desired maximum length of closed walks that we are interested in a signed network. Note that,

$$K(\beta,k) \approx \begin{cases} \dfrac{\sum_{i=1}^{n}\left(\frac{\lambda_i(G)^3}{1-\beta\lambda_i(G)}\right)}{\sum_{i=1}^{n}\left(\frac{\lambda_i(|G|)^3}{1-\beta\lambda_i(|G|)}\right)}, & \text{if } 0 < \beta < 1/\rho(|G|), k \to \infty \\ \\ \dfrac{\sum_{i=1}^{n}\lambda_i(G)^3}{\sum_{i=1}^{n}\lambda_i(|G|)^3} = K_{\triangle}, & \text{if } \beta \to 0. \end{cases}$$

Consequently, we define a measure of lack of balance of a signed network as

$$U(\beta,k) = \frac{1-K(\beta,k)}{1+K(\beta,k)}. \tag{3.5}$$

In particular,

$$U_{\triangle} = \frac{1-K_{\triangle}}{1+K_{\triangle}}. \tag{3.6}$$

3.2.1 Values of parameters β, k

Since the proposed measure of lack of structural balance $U(\beta,k)$ depends on the parameters β and k, it is natural to ask which values of these parameters provide an optimal choice for computing $U(\beta,k)$. However, this a difficult question to answer as it depends on the structure of the signed network. Indeed, it is plausible to investigate the behavior of $U(\beta,k)$ numerically for a given network when one of the parameters varies and the other one is fixed.

We recall that the parameter β is called the Katz parameter when the Katz index is used to find similarity of a pair of vertices in unsigned networks [Katz, 1953]. However, there is no agreed mechanism for selection of this parameter and the proposed value of $\beta = (1-e^{-\lambda_1})/\lambda_1$ as it is investigated in [Aprahamian et al., 2015] in view of centrality vectors, where λ_1 is the largest eigenvalue of the network. Indeed, when λ_1 is large it is computationally a challenging problem to compute $e^{-\lambda_1}$ efficiently. For our numerical simulation on real world signed networks we consider $k \to \infty$ and plot the graph of $U(\beta,k)$ when the value of β gradually increases from 0. It is interesting to notice that $U(\beta,k)$ increases exponentially as β increases which we discuss in the next subsection. A special attention is also given to

$$\beta \in \left\{ \beta_1 = \frac{0.85}{\rho(|G|)}, \beta_2 = \frac{1}{2\rho(|G|)}, \beta_3 = \frac{1}{||A||_\infty + 1} \right\},$$

where these values are considered in different contexts in studying centrality of vertices in unsigned networks [Benzi and Klymko, 2013; Amancio et al., 2012; Foster et al., 2001]. Further, by setting $\beta = \beta_1, \beta_2, \beta_3$ we plot $U(\beta,k)$ and observe that when k gradually increases from 1 to 50, the value of $U(\beta,k)$ increases in the beginning and after a certain critical value of k, $U(\beta,k)$ stabilizes for Epinions, Slashdot, and WikiElection which are real world large signed networks.

After the experiments on real world large networks that are discussed in the next subsection, we observe that $U(\beta,k)$ is a monotonically increasing function with respect to both β and k when one of them is fixed and the other one varies. Thus, finally we conclude that it would be close to impossible to come up with an absolute measure which can provide a concrete idea of structural balance of a signed network but any proposed measure can help to compare the degree of balance or unbalance between two signed networks of approximately same size.

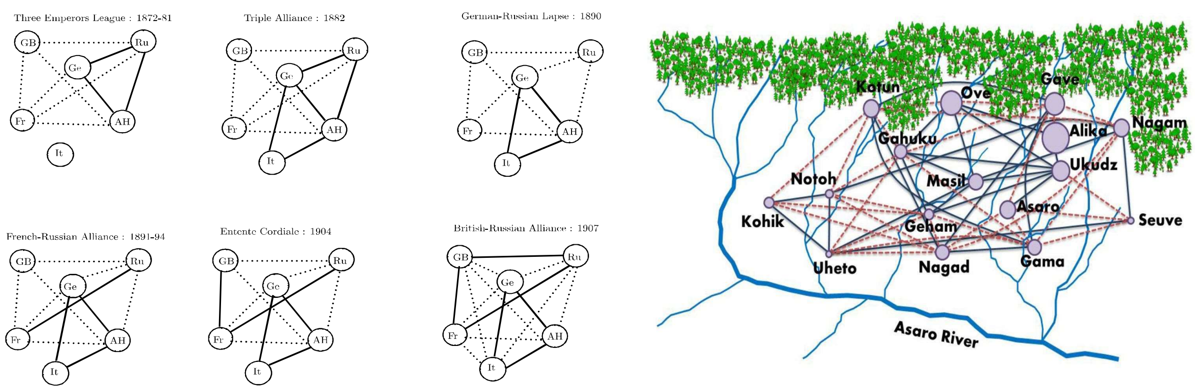


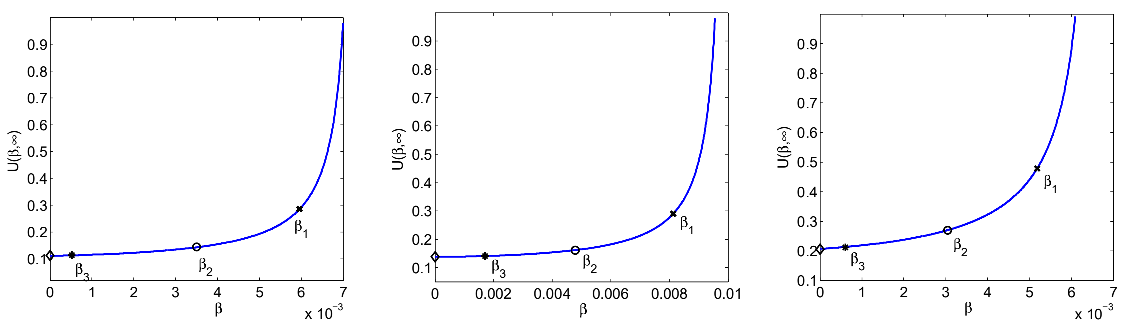
Figure 3.1 : (Left) Evolution of the balance among the six major players of the World War I at different time periods. Solid lines account for alliances and broken lines represent enmities. GB: Great Britain; Ru:Russia; Ge: Germany; Fr: France; AH: Austro-Hungarian Empire; It: Italy. (Right) Balance among the subtribes in the highlands of New Guinea [Hage, 1979]. Solid dark blue lines are for alliance (Rova) relations and red dashed lines are for antagonistic (Hina) relations.

3.2.2 Lack of balance in random and real world signed networks

In order to analyze the performance of the proposed measure, we consider a few real world networks which are also used to study the performance of U in [Estrada and Benzi, 2014]. The small scale networks which are used in our study include the networks which represent the evolution of the relations between the major players in the World War I (WWI) (Figure 3.1) [Antal et al., 2006] and the networks which provide the Gahuku-Gama subtribe system in the Eastern Central Highlands of New Guinea figure (Figure 3.1) [Hage, 1979]. For large signed networks, as mentioned earlier, we consider Epinions: a trust-distrust network among users of the product review site Epinions[Guha et al., 2004], Slashdot: a friend-foe network in the technological news site Slashdot[Lampe et al., 2007], and WikiElection: a network representing the votes for the election of administrators in Wikipedia [Burke and Kraut, 2008].

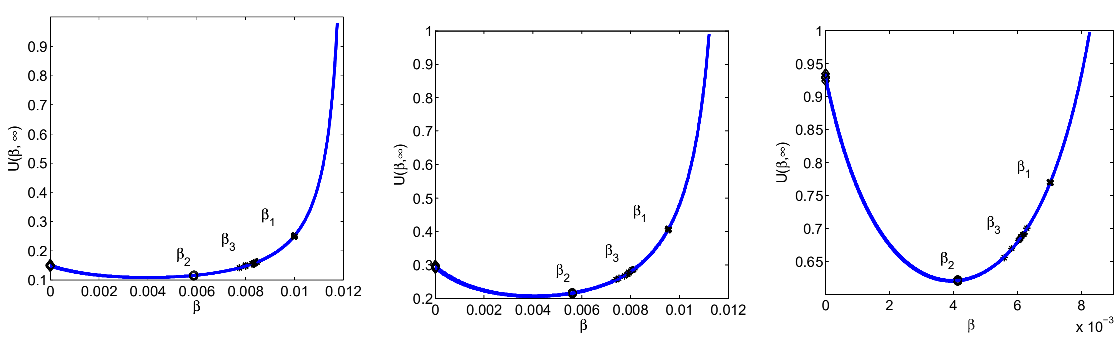
Table 3.2 : Degree of lack of balance of networks for different parameters.

Networks	U	$U(\beta_1,\infty)$	$U(\beta_2,\infty)$	$U(\beta_3,\infty)$	$U_\triangle$
Three Emperor's League	0.2043	0.5183	0.3573	0.4192	0.4000
Triple Alliance	0.1771	0.4269	0.2855	0.3070	0.3333
German-Russian Lapse	0.1798	0.5148	0.3910	0.3919	0.6667
French-Russian Alliance	0.1850	0.4507	0.2937	0.3202	0.3333
Entente Cordiale	0.1771	0.4269	0.2855	0.3070	0.3333
British-Russian Alliance	0	0	0	0	0
Subtribes in the highlands of New Guinea	0.4674	0.4841	0.2498	0.3496	0.1935
Wikielection	1	0.2859	0.1446	0.1147	0.1126
Slashdot	1	0.2902	0.1618	0.1410	0.1382
Epinions	1	0.4782	0.2691	0.2118	0.2051
RN-I	1	0.2501	0.1148	0.1531	0.1487
RN-II	1	0.4044	0.2141	0.2817	0.2912
RN-III	1	0.7686	0.6192	0.6884	0.9270



(a) (b) (c)

Figure 3.2 : Variation of degree of unbalance $U(\beta,\infty)$ w.r.t β (a) WK (b) SD (c) EPN



(a) (b) (c)

Figure 3.3 : Variation of degree of unbalance $U(\beta,\infty)$ of random networks w.r.t β (a) RN-I (b) RN-II (c) RN-III

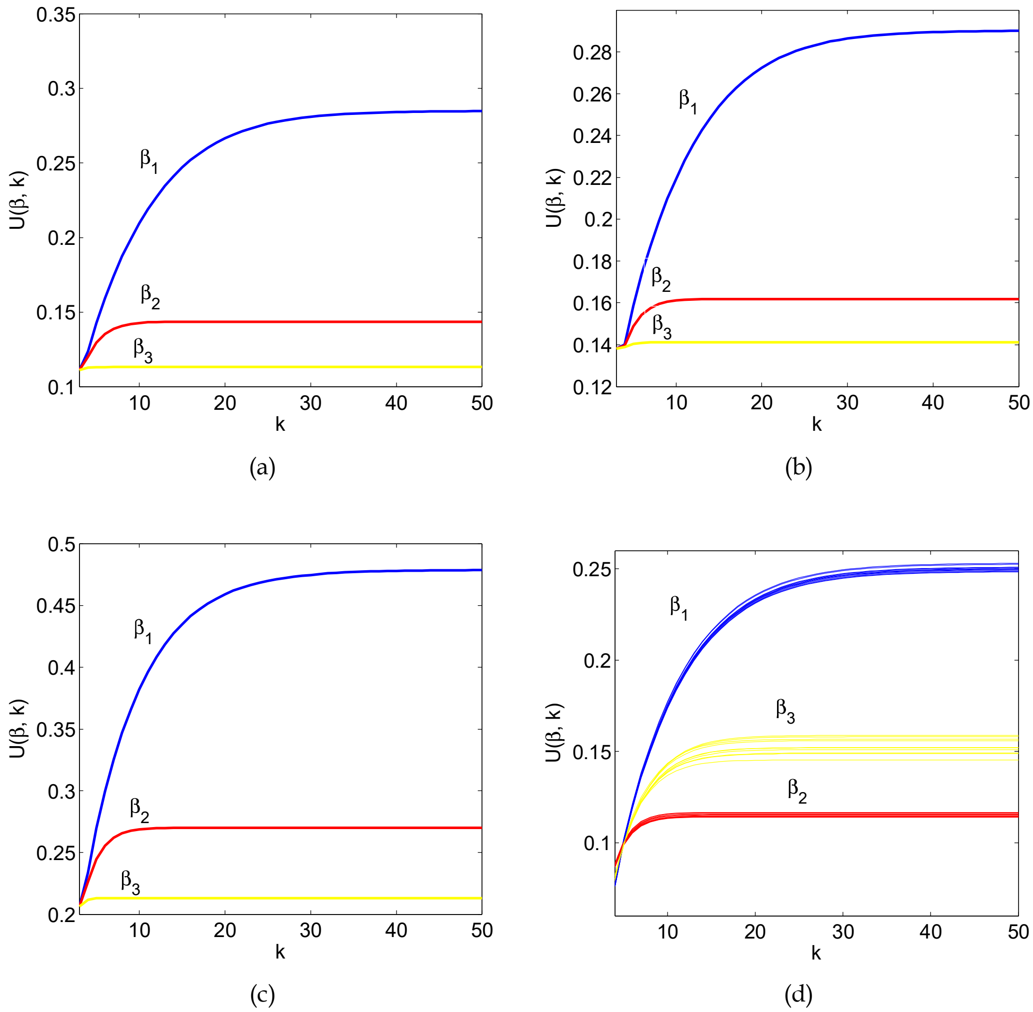


Figure 3.4 : Variation of degree of Unbalance w.r.t length of closed walks for a given β (a) WK (b) SD (c) EPN (d) RN-I

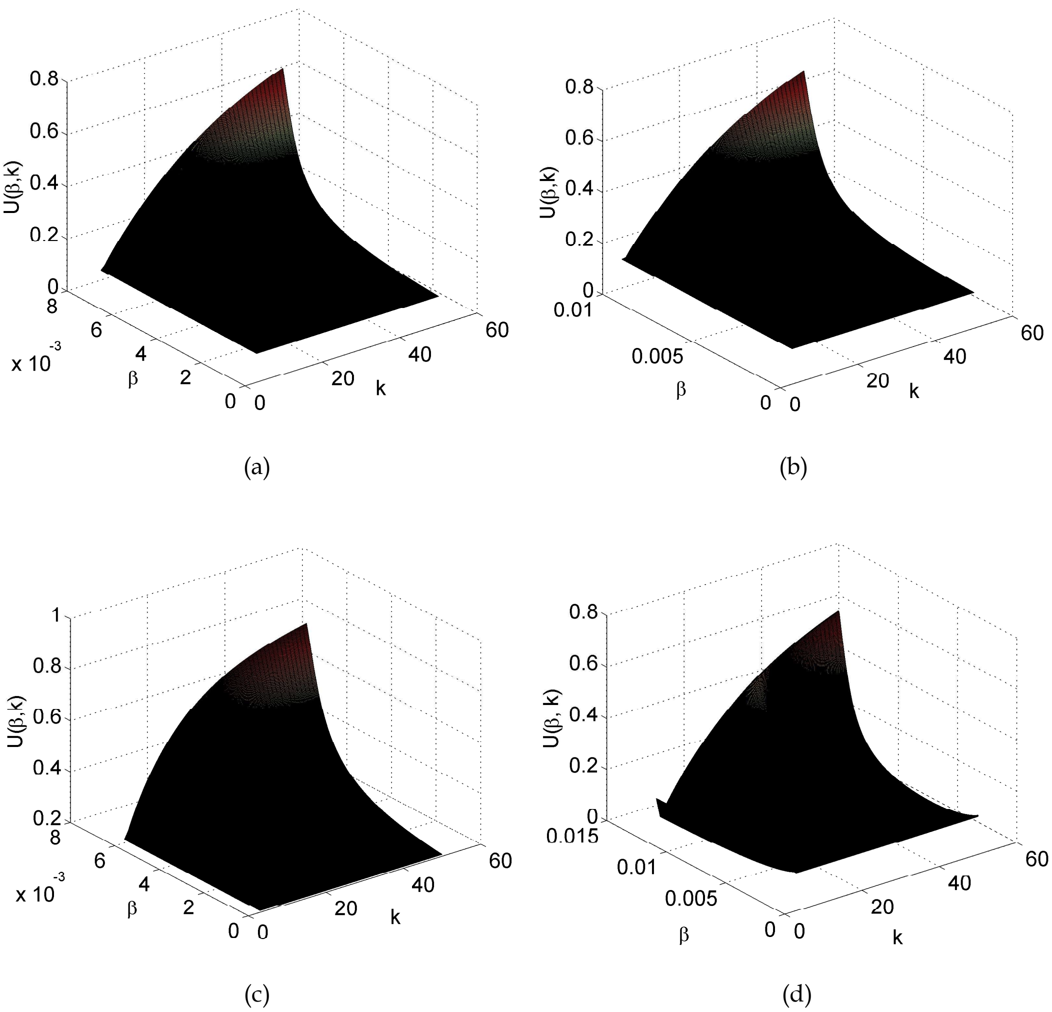


Figure 3.5 : Variation of degree of Unbalance as a function of β, k (a) WK (b) SD (c) EPN (d) RN-I

We also consider the performance of the proposed measure for random signed networks. We generate an ensemble of 10 random signed networks in which an edge between a pair of vertices exists with a positive sign with probability p, the negative sign with probability q and no edge between them with probability $1-p-q$ as described in [El Maftouhi et al., 2012]. We generate three random signed networks RN-I ($p = 0.01, q = 0.0005$), RN-II ($p = 0.01, q = 0.001$), and RN-III ($p = 0.01, q = 0.005$). Obviously, it produces random signed networks having more positive triangles than negative triangles, a phenomenon that occurs in real signed networks. The statistical details of such a network are provided in Table 3.1 by considering the average values of the parameters for each of the ensembles.

(a) *Small networks:* The networks associated with World War I (WWI), as observed in [Estrada and Benzi, 2014], the relations between the countries depicted in the corresponding signed graphs evolve and the structural balance increases gradually starting from 1872-81 to 1907. The same results are achieved by using the proposed measure. However, the degree of lack of balance is more in all the networks in the proposed measure compared to their lack of balance provided in [Estrada and Benzi, 2014] as follows from Table 3.2. In the network of the subtribes in the highlands of New Guinea, $U \approx U(\beta_1, \infty)$ and the value indicates that the network is almost far from structural balance, however, for $\beta = \beta_2, \beta_3$ the value of $U(\beta, \infty)$ show that the network is fairly balanced like the value of $U_{\triangle}$, see Table 3.2.

(b) *Large networks:* In contrast to the observation in [Estrada and Benzi, 2014] that Epinions, Slashdot, and WikiElection are totally unbalanced networks, using the proposed measure we find out that it is not a rational observation. In fact, as it is shown in Table 3.2, these networks are quite structurally balanced as also concluded in [Facchetti et al., 2011] where in their proposed method they give significantly more weight to the contributions of triads to the degree of balance in these networks.

(c) *Random networks:* The degree of unbalance for different values of β of three random networks RN-I, RN-II, RN-III having an approximately same number of vertices and edges are given in Table 3.2. We consider an ensemble of 10 random networks for each pair of parameter values p, q. From equation (3.1), the method proposed in [Estrada and Benzi, 2014] determines that these random networks are 100% unbalance, hence we can not distinguish these random networks based on their lack of balance. Thus, similar to real world networks, it gives a flawed degree of unbalance of random networks. Whereas, our proposed method gives different values of their degree of unbalance depending on the values of the parameters in the metric. It is to note that, like real networks, now random networks can also be distinguished based on their degree of unbalance for some given value of parameter β. This makes our proposed measure a better quantification of the degree of unbalance in signed networks.

We emphasize that, as argued in Subsection 3.1.1, the claim in [Estrada and Benzi, 2014] regarding the degree of lack of balance is biased towards unbalance for both real empirical networks and the random networks considered in the chapter. Whereas, the proposed measure not only provides a reasonable way to quantify the degree of lack of balance it also enables to distinguish signed networks based on their degree of unbalance. It is to be noted that for small networks like networks associated with WW1, a minor change of a number of signs of edges may have a significant impact on the degree of unbalance due to the formulation of measure using eigenvalues and other algebraic properties, and hence degree of unbalance by counting weighted cycles should be recommended. For the case of large networks where counting weighted cycles is a hard problem [West et al., 2001], measures using eigenvalues and other algebraic properties are useful.

3.2.3 Variation with parameters

1. *Varying β when $k \to \infty$* : In Figure 3.2 we show the values of $U(\beta,k)$ as β grows from 0 to $1/\rho(|G|)$ keeping $k \to \infty$ fixed. The results establish that the degree of unbalance grows exponentially up to almost 1 as β increases in this case and this happens in all the three large empirical networks. Thus when $k \to \infty$ it would not be rational to declare that a network is a balance/unbalance for any value of β. Indeed for a fixed value of β we can use this metric to compare the degree of balance for a given collection of networks. Here, setting $\beta = \beta_1$, Epinions is the most unbalance while Wikipedia is least. In Figure 3.3 we show the corresponding results for the random signed networks whose details are mentioned above. Observe that RN-I, RN-II show the same trend alike real empirical networks except the fact that $U(\beta_3,k) > U(\beta_2,k)$. In RN-III, the numbers of balance and unbalance triangles are almost equal and hence it makes RN-III more unbalance compare to RN-I and RN-II.

2. *Varying k when β is fixed:* In Figure 3.4 we show the performance of the metric $U(\beta,k)$ when the length of closed walks (k) varies and $\beta \in \{\beta_1,\beta_2,\beta_3\}$ is fixed. It is interesting to observe that the growth of $U(\beta,k)$ becomes close to zero, that is, $U(\beta,k)$ becomes almost constant after some threshold value of k in all the networks. We remark that this is indeed not surprising since β is the attenuation factor in Katz measures, and β^l weighted the contribution of the cycle of different length l in the measure, such that a small β would make the contribution from long cycle vanishing. Hence the smaller the β, the smaller the value of k that the measure $U(\beta,k)$ saturates. For example, in Wikipedia, the threshold values of k are $30, 10$ and 3 for $\beta = \beta_1, \beta_2$ and β_3 respectively. Thus we can conclude from these numerical results that we need not consider all values of k up to ∞ in the formula of $U(\beta,k)$ but a suitable finite value of k can decide the degree of unbalance/balance of a network if this measure is used. Also observe that for the random signed networks $U(\beta,k)$ become almost constant after some threshold value of k.

3. *Varying both β and k* : In Figure 4.2 we plot the surface $U(\beta,k)$ when β varies from 0 to $1/\rho(|G|)$ and k is from 3 to 50. Observe that $U(\beta,k)$ increases as both k and β increase and for the lower value of both k and β, the degree of lack of balance is very low. Of course, any fixed value of the pair (β,k) can be used to compare the lack of balance using the metric $U(\beta,k)$. Indeed, it would be an interesting problem to find an optimal choice of values for both β and k to compare the lack of balance for two given networks.

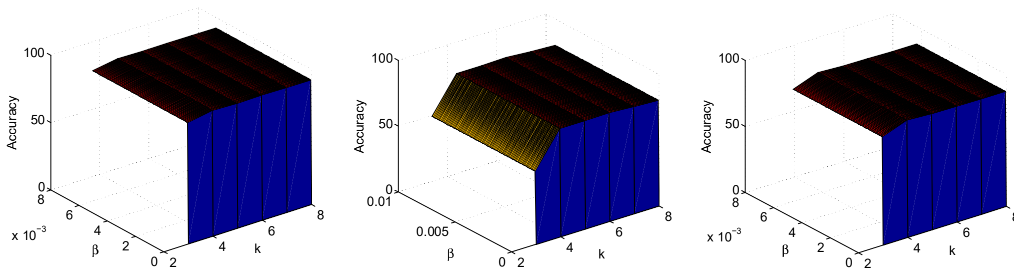
3.3 SIGN PREDICTION IN SIGNED NETWORKS

From Section 2.3 the predicted sign of an edge between u and v is defined by

$$sign\left\{\mu(G^{-(uv)}) - \mu(G^{+(uv)})\right\} = sign\left(\sum_{t=3}^{k} \beta^t A_{uv}^{t-1}\right). \tag{3.7}$$

In particular, by considering the weight factor β^t as β^{t-1}, the Equation (3.7) becomes the Katz prediction rule given by $sign\left\{(I-\beta A)^{-1} - I - \beta A\right\}_{uv} := (P(\beta,\infty))_{uv}$, $0 \leq \beta < 1/\rho(G)$ which essentially is the instrument for the parametrized measure for the lack of balance proposed in Section 3.1.

On the other hand the use of exponential of the adjacency matrix for the definition of walk based measure for lack of balance in [Estrada and Benzi, 2014] provides the sign prediction rule as $\{e^A - A - I\}_{uv} := (P)_{uv}$ for a pair of vertices u, v. We compare the performances of both the prediction rules P and $P(\beta,\infty), \beta \in \{\beta_1,\beta_2,\beta_3\}$ to predict sign of an edge with unknown sign in the real world signed networks Epinions, Wikielection and Slashdot. We proceed as follows. First, we remove 10% of edges randomly from these networks and denote it by E_{test}. Consider the resultant



(a) (b) (c)

Figure 3.6 : Accuracy of sign prediction as a function of β, k (a) WK (b) SD (c) EPN

network and predict the signs of deleted edges which are in E_{test} by using these prediction rules. The accuracy of the prediction rule is considered to be the % of successful sign predictions of edges. Here we do the experiment of sign predictions by using 10-fold cross validation due to the randomness in selecting the edges to be deleted.

Note that a large majority of edges are positive in the three empirical networks and it implies that one can achieve the accuracy equal to the fraction of positive edges, on average, even if guessing the sign of focal edge as positive. Thus we consider the ratio of a number of positive edges to the total number of edges as the baseline to compare accuracy. Then the computed average baselines for 10-fold cross validation after removal of 10 % edges are 91.8%, 73.8%, and 84.9% for Wikielection, Slashdot, and Epinions, respectively. The % of successfully predicted signs using different β values are given in Table 3.3.

Further note that using the known signs of 90% edges in order to predict signs of 10% of edges may not be very realistic since in many cases, for instance, signs of the known 90% of edges need not be exact due to an unreliable source. Thus we examine the performance of the proposed sign prediction rule after sparsifying the existing networks after the removal of 50% of its edges randomly. Then the average baselines for these sparsified networks after further removing 10% edges are 88.67%, 71.89%, and 83.17% for Wikielection, Slashdot, and Epinions, respectively in 10-fold cross validation. In Table 3.4 we present the accuracy of sign prediction for both the measures $P(\beta, \infty)$ and P.

As it follows from Table 3.3 and Table 3.4 that the prediction results using the proposed method are better than the baselines as well as that for P. Results on Slashdot are better than rest of the networks. Among β values, $P(\beta_1, k)$ has highest prediction accuracy.

Table 3.3 : % of successfully predicted sign using different β values and P

Networks	$P(\beta_1, \infty)$	$P(\beta_2, \infty)$	$P(\beta_3, \infty)$	P
Wikielection	92.34	92.16	92.10	91.60
Slashdot	82.30	81.62	81.24	73.60
Epinions	87.62	87.12	87.34	85.2

A pertinent question about the Katz prediction rule defined in Equation (3.7) is whether the role of longer cycles in the prediction of signs of edges is important or not. Intuitively, when cycles are counted of all lengths up to infinity many edges get repeated for larger cycles. Hence

Table 3.4 : % of successfully predicted sign using different β values and P for Sparse dataset

Networks	$P(\beta_1,\infty)$	$P(\beta_2,\infty)$	$P(\beta_3,\infty)$	P
Wikielection	90.26	89.88	89.60	90.10
Slashdot	78.58	78.02	77.46	74.20
Epinions	85.86	85.54	84.86	84.20

we pose the following question. What is the optimal value of the parameter k in (3.7) in order to successfully predict the sign of an unknown edge for a given signed network?

Note that the Figure 3.6 demonstrates the prediction accuracy as a function of β and k for the empirical networks. Observe that, in general, for all values of β prediction accuracy is best in the range $k = 4$ to $k = 6$ and it is worst for $k = 3$. For the walks of length greater than 6, accuracy first slightly decreases and then becomes almost constant. Thus the prediction accuracy surface recommends that we need not use the longer walks for prediction since shorter walks give better results. Also, observe that in general for the larger value of β accuracy is slightly better than that for a small value of β. Finally, it becomes an interesting fact that the effects of these two parameters are opposite; increasing β value puts more weight on longer walks whereas decreasing k value cut the contribution of longer walks. Finally we conclude that the longer closed walks are not much important for sign prediction neither are the triads; in other words, shorter closed walks of lengths $4,5,6$ are important factors in sign predictions in signed social networks.

3.4 CONCLUSION

In this chapter, we have proposed a method based on weighted closed walks to measure the degree of unbalance in signed networks. The spectra of the signed network have a key role in developing the measure of the degree of unbalance. The proposed measure can be used to distinguish two sign networks based upon their degree of unbalance. Later we use this measure for prediction of signs of edges in the real world signed networks and we observe that the closed walk of shorter length is efficient in sign prediction.