Journal of Statistical Mechanics:
Theory and Experiment

PAPER: DISORDERED SYSTEMS, CLASSICAL AND QUANTUM

Measuring the balance of signed networks and its application to sign prediction

To cite this article: Ranveer Singh and Bibhas Adhikari *J. Stat. Mech.* (2017) 063302

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

PAPER: Disordered systems, classical and quantum

Measuring the balance of signed networks and its application to sign prediction

Ranveer Singh[1] **and Bibhas Adhikari**[2]

[1] Centre for System Science, Indian Institute of Technology Jodhpur, Ratanada, Jodhpur, Rajasthan, India
[2] Department of Mathematics, Indian Institute of Technology Kharagpur, Kharagpur, West Bengal, India
E-mail: pg201283008iitj.ac.in and bibhasmaths.iitkgp.ernet.in

Received 14 January 2017
Accepted for publication 15 May 2017
Published 28 June 2017

Online at stacks.iop.org/JSTAT/2017/063302
https://doi.org/10.1088/1742-5468/aa73ef

Abstract. We propose a parametrized walk-based measure for the lack of balance in signed networks inspired by the Katz measure of similarity of two vertices in a network. We show that the performance of the proposed measure is marginally better than a recently proposed walk-based measure of the lack of balance for an undirected version of the real-world signed networks: Epinions, Slashdot and WikiElection. The proposed measure can be used to distinguish signed social networks on the basis of their degree of lack of balance. We also establish that cycles of shorter lengths can predict the sign of an edge in these signed networks better than the longer cycles by using the Katz prediction rule.

Keywords: communication, supply and information networks, critical phenomena of socio-economic systems, random graphs, networks, socio-economic networks

 1742-5468/17/063302+16$33.00

Contents

1. Introduction

Despite the existence of myriad and complex relationships between social entities in the real world, the main focus in the area of social networks has, until very recently, been given to friendships. A network is considered as a social network when the vertices represent social entities and a social relationship between two vertices is represented by a link in the network. In his seminal work on the analysis of balance cognitive units, the Austrian psychologist, Fritz Heider, distinguished between two major types of relations [1]. One is concerned with the relationship of love or likings, and the other one is about the relation of hate or disliking. The concept of the balanced state of a relationship between three social entities was introduced by considering certain combinations of these relations. Later, in 1956, the American mathematician, Frank Harary, modeled the cognitive structure of balance which is consistent with Heider's concept of balance by introducing the concept of signed graphs [2, 3]. In a signed graph, the vertices represent individuals and a positive link (link with a positive sign) between two vertices reflects the existence of a liking relationship, whereas, a negative link (link with a negative sign) represents disliking.

In the context of understanding relationships of a person with others, the concept of balanced state was introduced by Heider in [1]. Besides, the attitude of a person towards other persons was assessed by two types of triples which are involved in signed relations: those involving three individuals and those of two individuals and a social object such as a belief. Later, Cartwright and Harary [2] generalized and extended this concept in the language of signed graphs in 1956, where such a graph structure was proposed in order to obtain a mathematical model of the balanced cognitive unit. In their study of structural balance, the concept of balance for a triad was proposed. A triad, which is a completely connected graph on three vertices, is called balanced when the

J. Stat. Mech. (2017) 063302

product of signs of its edges is positive. Otherwise, a triad is called unbalanced. When this notion of balanced cycles is applied to signed networks it leads to the following theorems for structural balance to be obtained. As an outcome of the study of Heider, it is believed that a signed social network evolves towards a balanced state otherwise a state of unbalance will produce tension.

Theorem 1.1 ([3]). *For a balanced signed network, either all its edges are positive or the vertices can be partitioned into two subsets such that each positive edge joins vertices in the same subset and each negative edge joins vertices in different subsets.*

Theorem 1.2 ([3]). *A signed network is balanced if and only if for each pair of distinct vertices u and v all paths connecting u and v have the same sign.*

Theorem 1.3 ([3]). *A signed network is balanced if and only if all cycles are balanced.*

It may be noted that the theorems mentioned above can be used as criteria to determine whether a given signed graph is balanced. Nonetheless, for substantially large networks, it could be a computationally challenging task to verify these criteria, especially in dynamic networks. Further, if a very small fraction of cycles in a large signed network is unbalanced it would be unfair to call the entire network unbalanced. It is also shown in [4, 5] that the undirected versions of some real-world signed social networks are not structurally balanced. Indeed, Epinions: a trust-distrust network among users of the product review site Epinions [6, 7], Slashdot: a friend–foe network in the technological news site Slashdot [7, 8], and WikiElection: a network representing the votes of the election of administrators in Wikipedia [7, 9] are all unbalanced real networks. These observations trigger off the following question: what level of balance exists in these networks? This calls for new criteria to measure the balance of a signed networks, especially for large signed networks.

We recall that several metrics are proposed in the literature for measuring structural balance. For example, closed cycle-based methods are proposed in [5, 2]. One of such measures which is used in model conflict dynamics [10] is given by the ratio of the number of signed to unsigned triangles in a signed network as follows

$$K_\triangle = \frac{\mathrm{trace}(A^3)}{\mathrm{trace}(|A|^3)} \tag{1}$$

where A denotes the adjacency matrix of the sigend network and $|A|$ is the adjacency matrix of the unsigned network constructed from the signed network by replacing all negative edges into positive edges. Obviously, $-1 \leqslant K_\triangle \leqslant 1$. Indeed, $K_\triangle = -1$ when all the triangles are unbalanced and $K_\triangle = 1$ when all triangles are balanced in the signed network. We mention that (1) is a special case of the relative m balance which is defined as the ratio of the number of positive cycles of length at most m to the total number of cycles of length at most m introduced by Norman and Roberts in 1978. The relative balance which was proposed in [11] is given by

$$K_m = \frac{\sum_{m\geqslant 3} f(m) X_m^+}{\sum_{m\geqslant 3} f(m)(X_m^+ + X_m^-)}, \tag{2}$$

where X_m^+ (X_m^-) denotes the number of positive (negative) cycles of length m, and $f(m)$ is a monotonically decreasing function that weights the relative importance of cycles

J. Stat. Mech. (2017) 063302

of length m. Recently, a closed walk-based method for measuring the balance of signed networks is developed and it concludes that signed social networks are highly unbalanced [5, 12, 13].

In this paper, we first show that the walk-based metric introduced in [5] 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 formlation of the metric proposed in [5]. Thus we introduce a new parametrized 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 [5], 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 paper also contradicts the claim made in [5] that large signed random networks are 100% unbalanced.[3]

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 the same number of nodes and links generated by using different values of q (the probability for negative links) [25], while the new measure gives different values of the 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 ramdom networks of approximately the same size.

We employ the proposed metric to three different random signed networks having approximately the same number of nodes and links generated by considering different probabilities for the creation of negative links and show that their lack of balance differ. Whereas, the measure proposed in [5] 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 deals with predicting the sign of an edge by using the signs of edges in the rest of the network. Here, we use the well-known Katz prediction rule as discussed in [14]. 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 paper.

2. Walk-based measure for the 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 [12, 13]. It would not be exaggerated to say that the idea of communicability has been exploited to introduce the walk-based measure for the lack of balance for signed networks in [5]. In this section, we briefly review the walk-based measure for balance which was introduced in [5] and provide a mathematical reasoning of how it led to the conclusion that real networks are poorly balanced. First, we recall the following preliminaries.

[3] We are grateful to one of the reviewers for this paragraph.

J. Stat. Mech. (2017) 063302

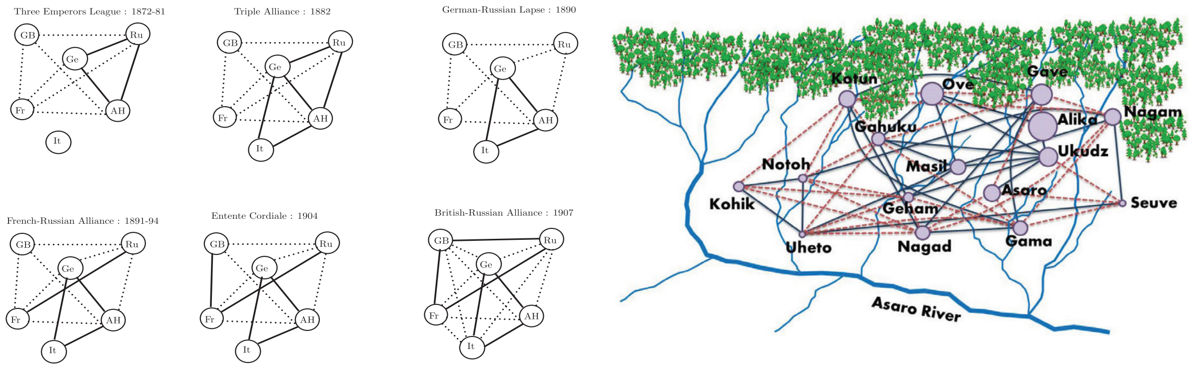


Figure 1. Top: evolution of the balance among the six major players of World War I at different time periods. Solid lines account for alliances and broken lines represent enmities. GB: Great Britain; Ru: Russia; Ge: Germany; Fr: France; AH: Austro-Hungarian Empire; It: Italy. Bottom: balance among the subtribes in the highlands of New Guinea [24]. Solid dark blue lines are for alliance (Rova) relations and red dashed lines are for antagonistic (Hina) relations.

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 link 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 $\mathrm{trace}(A^k)$. 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 the lack of balance in a signed network G on n vertices is defined by

$$U = \frac{1-K}{1+K}, \text{ where } K = \frac{\mathrm{trace}(e^A)}{\mathrm{trace}(e^{|A|})} = \frac{\sum_{j=1}^{n} \exp(\lambda_j(G))}{\sum_{j=1}^{n} \exp(\lambda_j(|G|))}, \tag{3}$$

$\lambda_j(G)$ and $\lambda_j(|G|)$, $j = 1, \ldots, n$ are eigenvalues of A and $|A|$ respectively, in ascending order [5]. Thus U can be interpreted as the ratio of weighted unbalanced to balanced closed walks. Note that, in calculating the lack of balance the weights of an m length walk

J. Stat. Mech. (2017) 063302

Table 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	92 238	7784	651 560	72 398	0.0039	142.7757	130.6673
Slashdot	9000	75 462	25 707	226 044	31 228	0.0025	104.53	95.9208
Epinions	8000	91 498	14 308	713 563	147 369	0.0033	164.0028	138.7156
RN-I	8000	319 760	15 979	171 625	25 528	0.0105	84.94	77.06
RN-II	8000	320 426	31 872	88 072	25 644	0.0110	89.0591	73.3225
RN-III	8000	319 180	159 953	148 767	137 909	0.0150	120.7704	42.8063
WWI	n	m^+	m^-	$\triangle^+$	$\triangle^-$	$\|A\|_\infty$	$\lambda_n(\|G\|)$	$\lambda_n(G)$
Three emperor's league	6	3	6	5	2	4	3.6458	3.1028
Triple alliance	6	5	6	6	2	5	3.8590	3.4163
German–Russian lapse	6	3	7	3	2	5	3.5141	3.0144
French–Russian alliance	6	4	7	6	2	5	3.8590	3.3743
Entente cordiale	6	5	6	6	2	5	3.8590	3.4163
British–Russian alliance	6	6	9	20	0	5	5	5

is assumed to be $1/m!$ which is a decreasing function of length. Note also that this measure (3) has a resemblance of the measure (2). It follows from the theorem 2.1 stated below that for balanced networks, that is, 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 $\sum_{j=1}^n \exp(\lambda_j(G)) \ll \sum_{j=1}^n \exp(\lambda_j(|G|))$.

Theorem 2.1 ([15]). *Matrices M and $|M|$ are isospectral if and only if graph corresponding to M, is cycle balanced.*

2.1. 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 2, [5]). This result has been justified by supposing 'the triads with only one negative link or with all three negative links have been found to be overrepresented in the three online networks' which is an observation made in [16]. However, it follows from the definition of K that

$$K = \frac{\sum_{j=1}^n \exp(\lambda_j(G))}{\sum_{j=1}^n \exp(\lambda_j(|G|))} = \frac{\exp(\lambda_n(G))}{\exp(\lambda_n(|G|))}\frac{\mathcal{O}(n)}{\mathcal{O}(n)} \tag{4}$$

and from table 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

J. Stat. Mech. (2017) 063302

of G. We mention that all the large signed networks including the random networks that are considered in this paper 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 $\rho(G) < \rho(|G|)$ (p 619, [17]) 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 (4).

Since, the connection between the 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 the lack of balance in signed networks.

2.2. Parameterized walk-based measure for the lack of balance

In this section, 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. The Katz measure is also valid for signed networks and gainfully used for link prediction problem in signed networks [18]. Note that, this is a resolvent centrality measure defined as

$$\sum_{l=0}^{\infty} \beta^l A_{ij}^l = \left((I - \beta A)^{-1}\right)_{ij}, \tag{5}$$

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 node i and $((I - \beta A)^{-1})_{ii}$ provides the weighted sum of the number of closed walks at the node 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 [14].

Inspecting the definition given in (3) it will be tempting to define a measure for the degree of balance of a signed network by using the 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 $\rho(A) \leqslant \rho(|A|)$ and n is the number of vertices in the network. However, observe that, however large the network is, the same links will be counted a large number of times for calculating the 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

$$K(\beta,k)=\frac{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^{l}A_{ii}^{l}}{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^{l}|A|_{ii}^{l}}=\frac{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^{l}\lambda_i(G)^{l}}{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^{l}\lambda_i(|G|)^{l}}$$
$$=\frac{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^{l-3}\lambda_i(G)^{l}}{\sum_{i=1}^{n}\sum_{l=3}^{k}\beta^{l-3}\lambda_i(|G|)^{l}}=\frac{\sum_{l=3}^{k}\beta^{l-3}\mathrm{Tr}(A^{l})}{\sum_{l=3}^{k}\beta^{l-3}\mathrm{Tr}(|A|^{l})}, \tag{6}$$

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,

$$K(\beta,k)\approx\begin{cases}\dfrac{\sum_{i=1}^{n}\left(\frac{\lambda_i(G)^3}{1-\beta\lambda_i(G)}\right)}{\sum_{i=1}^{n}\left(\frac{\lambda_i(|G|)^3}{1-\beta\lambda_i(|G|)}\right)}, & \text{if } 0<\beta<1/\rho(|G|), k\to\infty\\[2ex] \dfrac{\sum_{i=1}^{n}\lambda_i(G)^3}{\sum_{i=1}^{n}\lambda_i(|G|)^3}=K_{\triangle}, & \text{if } \beta\to 0.\end{cases}$$

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

$$U(\beta,k)=\frac{1-K(\beta,k)}{1+K(\beta,k)}. \tag{7}$$

In particular,

$$U_{\triangle}=\frac{1-K_{\triangle}}{1+K_{\triangle}}. \tag{8}$$

2.3. Values of parameters β, k

Since the proposed measure of the 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 is 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 [19]. However, there is no agreed mechanism for selection of this parameter and the proposed value of $\beta=(1-\mathrm{e}^{-\lambda_1})/\lambda_1$ as investigated in [20] 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 section. 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 [21–23]. 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

J. Stat. Mech. (2017) 063302

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 section, 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 the same size.

2.4. 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 [5]. The small scale networks which are used in our study include the networks which represent the evolution of the relations among the major players in World War I (WWI) (figure 1) [25] and the networks which provide the Gahuku–Gama subtribe system in the Eastern Central Highlands of New Guinea figure (figure 1) [24]. For large signed networks, as mentioned earlier, we consider Epinions: a trust-distrust network among users of the product review site Epinions [6], Slashdot: a friend–foe network in the technological news site Slashdot [8], and WikiElection: a network representing the votes for the election of administrators in Wikipedia [9].

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 positive sign with probability p, the negative sign with probability q and no edge between them with probability $1 - p - q$ as described in [11]. 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 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 [5], 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 the lack of balance is more in all the networks in the proposed measure compared to their lack of balance provided in [5] as follows from table 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 $U(\beta, \infty)$ shows that the network is fairly balanced like the value of $U_\triangle$, see table 2.

(b) *Large networks:* in contrast to the observation in [4, 5] 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 shown in table 2,

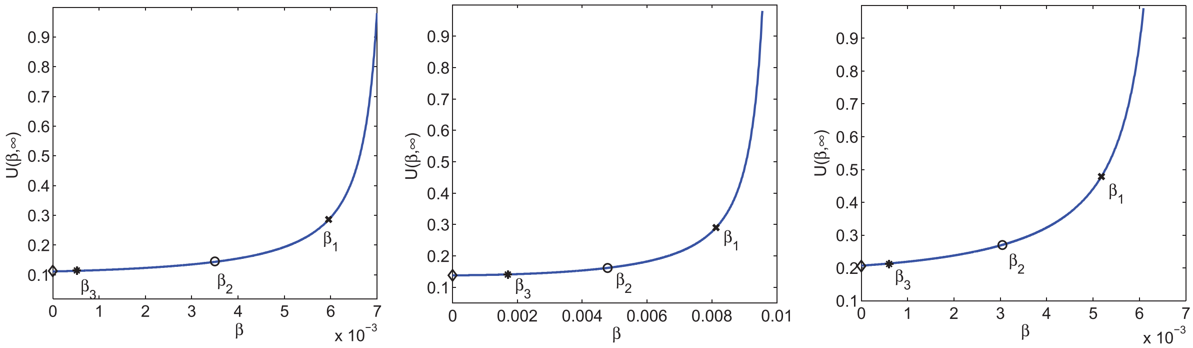
Table 2. Degree of the lack of balance of networks for different parameters.

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

these networks are quite structurally balanced as also concluded in [26] where in their proposed method they give significantly more weight to the contributions of triads to the degree of balance in these networks.

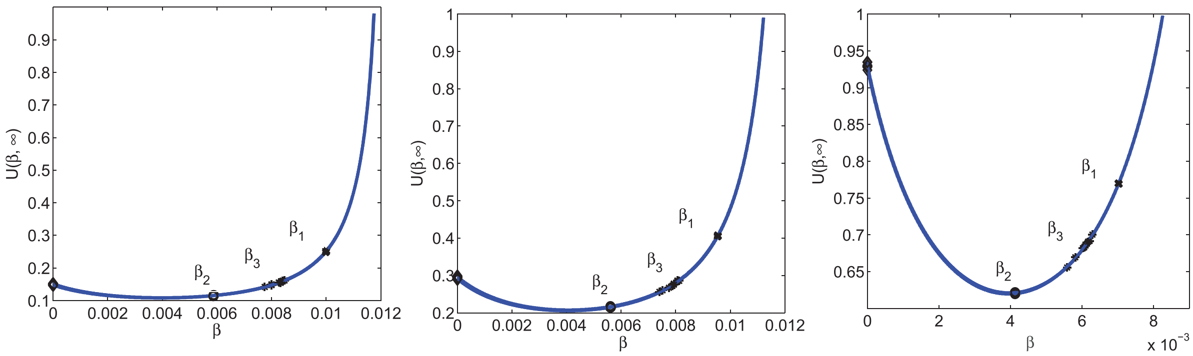
(c) *Random networks:* the degree of unbalance for different values of β of three random networks RN-I, RN-II, RN-III having approximately the same number of nodes and links are given in table 2. We consider an ensemble of 10 random networks for each pair of parameter values p, q. From equation (3), the method proposed in [4, 5] determines that these random networks are 100% unbalanced, 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 section 2.1, the claim in [4, 5] regarding the degree of the lack of balance is biased towards unbalance for both real empirical networks and the random networks considered in the paper. Whereas, the proposed measure not only provides a reasonable way to quantitate the degree of the lack of balance, it also enables signed networks to be distinguished based on their degree of unbalance. It is to be noted that for small networks like networks associated with WW1, a minor change of the number of signs of edges may have a significant impact on the degree on unbalance due to the formulation of measure using eigenvalues and other algebraic properties, and hence the degree of unbalance by counting weighted cycles should be recommended. For the case of large networks where counting weighted cycles is a difficult problem [27], measures using eigenvalues and other algebraic properties are useful.



(a) (b) (c)

Figure 2. Variation of the degree of unbalance $U(\beta,\infty)$ w.r.t β (a) WK, (b) SD, (c) EPN.



(a) (b) (c)

Figure 3. Variation of the degree of unbalance $U(\beta,\infty)$ of random networks w.r.t β (a) RN-I, (b) RN-II, (c) RN-III.

2.5. Variation with parameters

1. *Varying β when $k \to \infty$* : in figure 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 balanced/unbalanced 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 unbalanced while Wikipedia is least. In figure 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 similar to real empirical networks except the fact that $U(\beta_3,k) > U(\beta_2,k)$. In RN-III, the numbers of balanced and unbalanced triangles are almost equal and hence it makes RN-III more unbalanced compare to RN-I and RN-II.

2. *Varying k when β is fixed:* in figure 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, $U(\beta,k)$

J. Stat. Mech. (2017) 063302

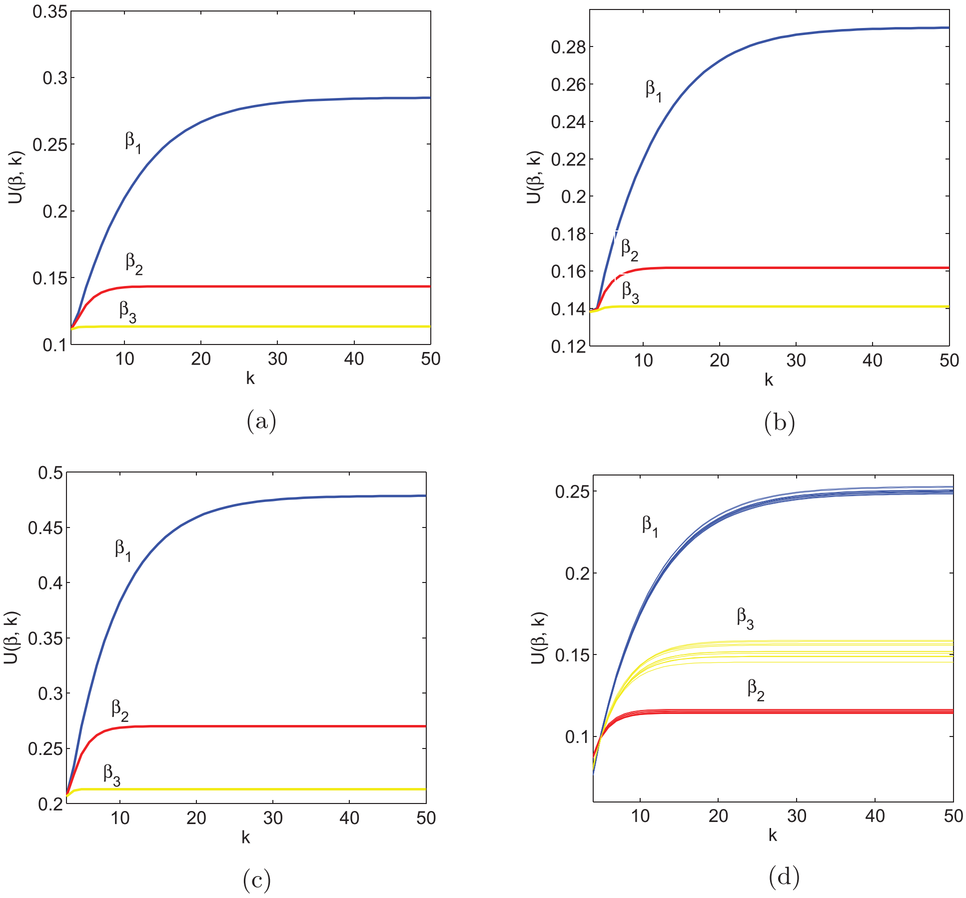


Figure 4. Variation of the degree of unbalance w.r.t length of the closed walks for a given β (a) WK, (b) SD, (c) EPN, (d) RN-I.

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 lengths 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 5 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 a lower value of both k and β, the degree of the 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.

J. Stat. Mech. (2017) 063302

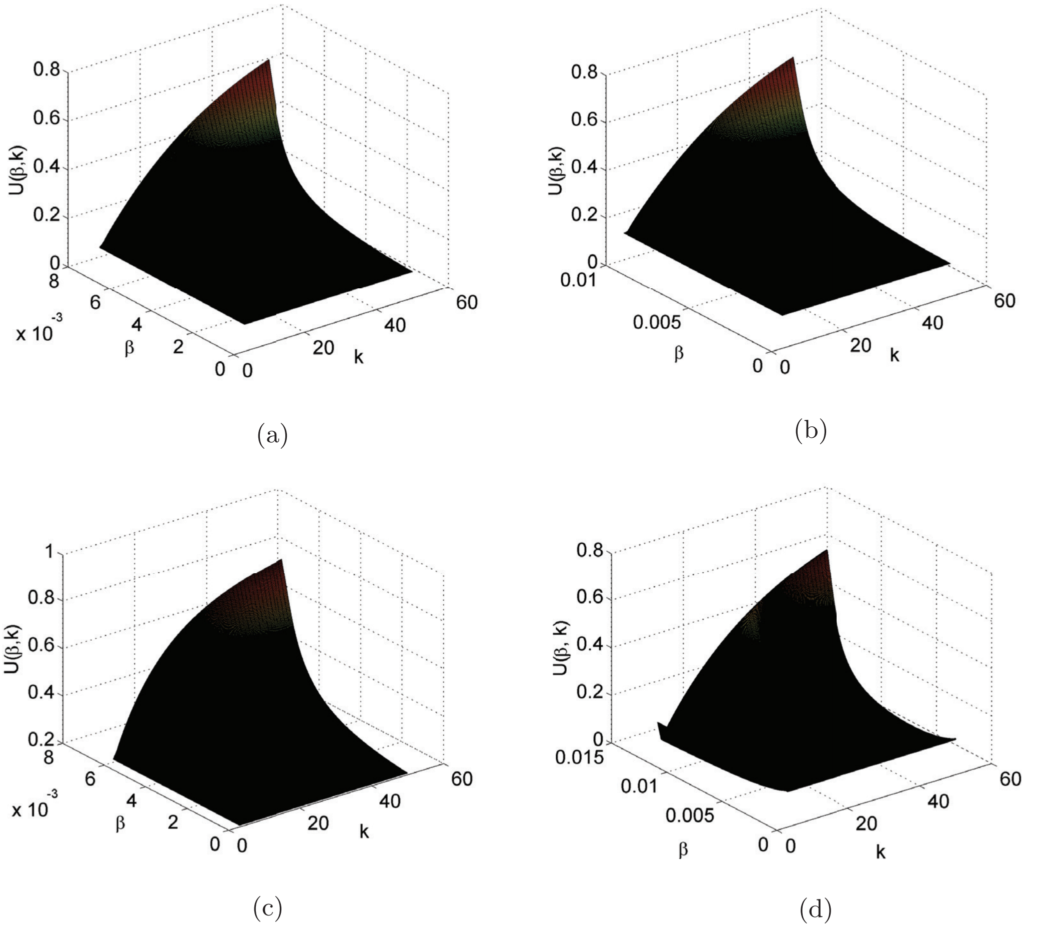


Figure 5. Variation of the degree of unbalance as a function of β, k (a) WK, (b) SD, (c) EPN, (d) RN-I.

3. Sign prediction in signed networks

The sign prediction problem deals with the inference of sign of an unknown link based upon observation of the entire signed network. The key idea for a prediction of sign of an edge is minimization of social unbalance [18] assuming that a signed network evolves towards balance. Since signs of edges contribute to calculating the degree of the lack of balance in a network, assigning the sign of an edge and keeping the signs of other edges fixed, it either minimizes or maximizes the social balance of a given network. Indeed, let us explain the idea of using balance for sign prediction as mentioned in [14] as follows. Consider two vertices u and v in the network such that $A_{uv} = 0$ and we have the task of predicting the sign of A_{uv}. First, add a positive edge between them and call the resulting augmented graph as $G^{+(uv)}$ and set $A_{uv} = 1$ in the original graph. Similarly, add a negative edge between u and v to construct the augmented graph $G^{-(uv)}$ and set $A_{uv} = -1$ in the original graph. Let $\mu(G)$ be the number of weighted unbalanced closed walks in graph G of length >2. Then the predicted sign of a link between u and v is defined by

$$\text{sign}\left\{\mu(G^{-(uv)}) - \mu(G^{+(uv)})\right\} = \text{sign}\left(\sum_{t=3}^{k} \beta^t A_{uv}^{t-1}\right). \tag{9}$$

Table 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

Table 4. % of successfully predicted signs using different β values and P for the 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

In particular, by considering the weight factor β^t as β^{t-1}, the equation (9) becomes the Katz prediction rule given by $\mathrm{sign}\{(I-\beta A)^{-1} - I - \beta A\}_{uv} := (P(\beta, \infty))_{uv}$, $0 \leqslant \beta < 1/\rho(G)$ which essentially is the instrument for the parametrized measure for the lack of balance proposed in section 2.

On the other hand, the use of the exponential of the adjacency matrix for the definition of a walk-based measure for the lack of balance in [5] 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 the 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 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 links are positive in the three empirical networks and it implies that one can achieve the accuracy equal to the fraction of positive links, on average, even if guessing the sign of focal link as positive. Thus we consider the ratio of the 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.

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 exisiting 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 4 we present the accuracy of sign prediction for both the measures $P(\beta, \infty)$ and P.

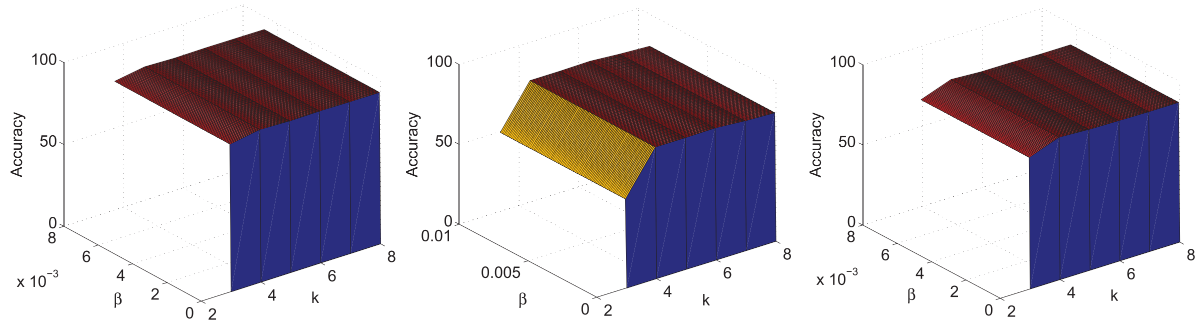


Figure 6. Accuracy of sign prediction as a function of β, k (a) WK, (b) SD, (c) EPN.

As follows from tables 3 and 4, the prediction results using the proposed method are better than the baselines as well as that for P. Results on Slashdot are better than the rest of the networks. Among β values, $P(\beta_1, k)$ has the highest prediction accuracy.

A pertinent question about the Katz prediction rule defined in equation (9) 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 we pose the following question. What is the optimal value of the parameter k in (9) in order to successfully predict the sign of an unknown edge for a given signed network?

Note that figure 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 walks with lengths 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 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 the β value puts more weight on longer walks whereas decreasing the k value cuts the contribution of longer walks. Finally, we conclue that the longer closed walks are not very important for sign prediction and 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.

4. Conclusions

In this paper, we have proposed a parametrized weighted closed walk-based method to measure the lack of balance in signed networks. We use the Katz prediction rule for prediction of signs of edges in real-world signed networks and we observe that cycles of shorter lengths can predict the sign of an edge compare to cycles of longer lengths.

Acknowledgment

The authors would like to thank the reviewers for their critical comments and suggestions which helped the authors to significantly improve the results and presentation of the paper.

References

[1] Heider F 1946 Attitudes and cognitive organization *J. Psychol.* **21** 107–12
[2] Cartwright D and Harary F 1956 Structural balance: a generalization of Heider's theory *Psychol. Rev.* **63** 277
[3] Harary F *et al* 1953 On the notion of balance of a signed graph *Michigan Math. J.* **2** 143–6
[4] Estrada E and Benzi M 2014 Are social networks really balanced? (arXiv:1406.2132)
[5] Estrada E and Benzi M 2014 Walk-based measure of balance in signed networks: detecting lack of balance in social networks *Phys. Rev.* E **90** 042802
[6] Guha R, Kumar R, Raghavan P and Tomkins A 2004 Propagation of trust and distrust *Proc. 13th Int. conf. on World Wide Web* (ACM) pp 403–12
[7] Leskovec J and Krevl A 2014 SNAP datasets: stanford large network dataset collection http://snap.stanford.edu/data
[8] Lampe C A, Johnston E and Resnick P 2007 Follow the reader: filtering comments on slashdot *Proc. SIGCHI Conf. on Human Factors in Computing Systems* (ACM) pp 1253–62
[9] Burke M and Kraut R 2008 Mopping up: modeling wikipedia promotion decisions *Proc. 2008 ACM Conf. on Computer Supported Cooperative Work* (ACM) pp 27–36
[10] Pelino V and Maimone F 2012 Towards a class of complex networks models for conflict dynamics (arXiv:1203.1394)
[11] El Maftouhi A, Manoussakis Y and Megalakaki O 2012 Balance in random signed graphs *Internet Math.* **8** 364–80
[12] Estrada E and Hatano N 2008 Communicability in complex networks *Phys. Rev.* E **77** 036111
[13] Crofts J J and Higham D J 2009 A weighted communicability measure applied to complex brain networks *J. R. Soc. Interface* **6** 411–4
[14] Chiang K-Y, Natarajan N, Tewari A and Dhillon I S 2011 Exploiting longer cycles for link prediction in signed networks *Proc. 20th ACM Int. Conf. on Information and Knowledge Management* (ACM) pp 1157–62
[15] Acharya B D 1980 Spectral criterion for cycle balance in networks *J. Graph Theory* **4** 1–11
[16] Leskovec J, Huttenlocher D and Kleinberg J 2010 Signed networks in social media *Proc. SIGCHI Conf. on Human Factors in Computing Systems (ACM)* pp 1361–70
[17] Meyer C D Matrix analysis and applied linear algebra
[18] Chiang K-Y, Hsieh C-J, Natarajan N, Dhillon I S and Tewari A 2014 Prediction and clustering in signed networks: a local to global perspective *J. Mach. Learn. Res.* **15** 1177–213
[19] Katz L 1953 A new status index derived from sociometric analysis *Psychometrika* **18** 39–43
[20] Aprahamian M, Higham D J and Higham N J 2015 Matching exponential-based and resolvent-based centrality measures *J. Complex Netw.* **4** 157–76
[21] Benzi M and Klymko C 2013 Total communicability as a centrality measure *J. Complex Netw.* **1** 124–49
[22] Amancio D R, Oliveira O N Jr and Costa L D F 2012 Structure–semantics interplay in complex networks and its effects on the predictability of similarity in texts *Physica* A **391** 4406–19
[23] Foster K C, Muth S Q, Potterat J J and Rothenberg R B 2001 A faster Katz status score algorithm *Comput. Math. Organ. Theory* **7** 275–85
[24] Hage P 1979 Graph theory as a structural model in cultural anthropology *Annu. Rev. Anthropol.* **8** 115–36
[25] Antal T, Krapivsky P L and Redner S 2006 Social balance on networks: the dynamics of friendship and enmity *Physica* D **224** 130–6
[26] Facchetti G, Iacono G and Altafini C 2011 Computing global structural balance in large-scale signed social networks *Proc. Natl Acad. Sci.* **108** 20953–8
[27] West D B 2001 *et al* 2001 *Introduction to Graph Theory* vol 2 (Upper Saddle River, NJ: Prentice hall)

J. Stat. Mech. (2017) 063302