Computing global structural balance in large-scale signed social networks

Giuseppe Facchetti [*], **Giovanni Iacono** [*], **and Claudio Altafini** [*]

[*] SISSA, International School for Advanced Studies, via Bonomea 265, 34136 Trieste, Italy.

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Structural balance theory affirms that signed social networks, i.e., graphs whose signed edges represent friendly/hostile interactions among individuals, tend to be organized so as to avoid conflictual situations, corresponding to cycles of negative parity. Using an algorithm for ground state calculation in large-scale Ising spin glasses, in this paper we compute the global level of balance of very large on-line social networks, and verify that currently available networks are indeed extremely balanced. This property is explainable in terms of the high degree of skewness of the sign distributions on the nodes of the graph. In particular, individuals linked by a large majority of negative edges create mostly "apparent disorder", rather than true "frustration".

social network theory | structural balance | Ising spin glass | combinatorial optimization

On-line social networks are examples of large-scale communities of interacting individuals in which local ties between users (friend, fan, colleague, but also friend/foe, trust/distrust, etc.) give rise to a complex, multidimensional web of aggregated social behavior [1, 2, 3, 4]. For such complex networks, the emergence of global properties from local interactions is an intriguing subject, so far investigated mostly at structural and topological level [5, 6, 7, 8, 2]. In social network theory [9, 10, 11], however, the content of the relationships is often even more important than their topology, and this calls for the development of novel analytical and computational tools, able to extrapolate content-related features out of the set of interactions of a social community. This is particularly challenging when, as in social networks retrieved from on-line media, the size of the community is very big, of the order of 10^5 individuals or higher.

A global property that has recently attracted some attention [12, 13, 1, 14] is determining the structural balance of a signed social network. Structural (or social) balance theory was first formulated by Heider [15] in order to understand the structure and origin of tensions and conflicts in a network of individuals whose mutual relationships are characterizable in terms of friendship and hostility. It was modeled in terms of signed graphs by Cartwright and Harary [16], see [11, 10] for an overview of the theory. The nodes of the graph represent users and the positive/negative edges their friendly/hostile relationships. It has been known for some time how to interpret structural balance on such networks [16]: the potential source of tensions are the cycles of the graph (i.e., the closed paths beginning and ending on the same node), notably those of negative sign (i.e., having an odd number of negative edges). It follows that the concept of balance is not related to the actual number of negative edges on the cycles but only to their parity, see Fig. 1 for an illustration on basic graphs. In particular, a signed graph is *exactly* balanced (i.e., tensions are completely absent) if and only if all its cycles are positive [16]. As such, structural balance is intrinsically a property of the network as a whole, not fragmentable into elementary subgraphs.

From a computational point of view, verifying if a signed undirected network is exactly balanced is an easy problem, which can be answered in polynomial time [17, 18, 19]. When instead a graph is not exactly balanced, one can compute a *distance to exact balance*, i.e., a measure of the amount of unbalance in the network. The most plausible distance is given by the *least number of edges* that must be dropped (or changed of sign) in order for the graph to become exactly balanced [16, 20, 21]. Computing this distance (called the "line index of balance" in [20, 21]) is an NP-hard problem, equivalent to a series of well-known problems, such as

- computing the ground state of a (nonplanar) Ising spin glass [22];
- solving a MAX-CUT problem [23, 24];
- solving a MAX-2XORSAT problem [25].

The equivalence with energy minimization of a spin glass has for example been highlighted recently in [26] (see also [27, 28]). In this context, a negative cycle is denoted a *frustration*, and frustrations are the trademark of complex energy landscapes, with many local minima whose structure and organization has been so far explored only in special cases. For instance, the case studied in [28], the fully connected graph, is unrealistic for real social networks, which usually have heterogeneous connectivity degrees. As a matter of fact, for what concerns the on-line signed social networks currently available (see Materials and Methods for a description), only an analysis of local, low-dimensional motifs has been carried out so far [12, 13]. This amounts essentially to the enumeration of the triangles and to their classification into frustrated / not frustrated, see [13, 1]. An alternative approach is taken in [18], where spectral properties of the Laplacian are investigated. For connected signed graphs, the magnitude of the smallest eigenvalue of the Laplacian is indicative of how unbalanced a network is, i.e., of how much frustration is encoded in the cycles of the networks.

Both approaches provide useful information in order to understand the structural balance of signed social networks. Yet, this information is partial and unsatisfactory. The small motif analysis, for example, only identifies the frustration on the smallest possible groups of interacting users, but overlooks more long-range conflicts associated to longer cycles (and larger communities), see Fig. 1(b) for an example. The spectral approach, on the contrary, gives an idea of the overall amount of frustration of the network, but does not provide any information on which relationships remain unbalanced. In terms of spin glasses, solving the problem globally and identifying the residual ineliminable tensions means computing the ground state(s) of a heterogeneous Ising spin glass, with bimodal bond distribution. For this class of problems, algorithms have been benchmarked only on graphs of size up to a few thousands nodes. An overview of the state of the art for spin glass ground state search is available in [29, 30], and for MAX-CUT in [31, 23, 24, 32].

We have recently introduced a novel efficient heuristic for ground state search on signed graphs. This method is presented in [33] in the context of monotonicity of biological networks. It is based on the application of equivalence transformations to the signed graph, called *gauge transformations* in the spin glass literature [34], or switching

Reserved for Publication Footnotes

equivalences in the theory of signed graphs [35]. The aim of these transformations is essentially to eliminate the so-called apparent disorder from the graph, while preserving the original frustration. Practically, this leads to the reduction of the number of negative edges in the graph, see Fig. 1(c), which simplifies the calculation of global balance. Our algorithm has proved capable of reaching very low energies on extremely large graphs. For two of the three signed networks reported in this study, in particular, our calculations are essentially exact (upper and lower bounds on the computed global balance differ by less than 1%).

Structural balance theory affirms that human societies tend to avoid tensions and conflictual relations. In a signed graph this translates into a level of balance higher than expected, given the network structure. The level of balance achieved by a network depends on the connectivity of the graph, on the percentage of negative edges and, most of all, on the distribution of these negative edges on the graph. Partial hints that social networks currently available are more balanced than expected are provided by both the small motif screening of [13, 1] and the spectral analysis of [18], and are confirmed by our analysis. In particular, on the social networks analyzed in this work we show that the chance that a null model has the level of balance of the true networks is essentially equal to zero. For all three networks, the level of balance turns out to be even less than the Shannon bound one obtains developing a rate-distortion theory for the null models [36, 25, 37]. What makes our signed networks so balanced is the skewed distribution of the signs of the edges on the users: users with a large majority of friends but also users with a large majority of enemies are not causing any significant frustration. In particular, when an individual is unanimously tagged as "enemy" by the other users, all the corresponding negative edges disappear if we apply a gauge transformation. As gauge transformations do not alter the sign of the cycles, it implies that these negative edges are indeed not contributing to the frustration but only to the apparent disorder, as in a so-called Mattis system [26]. In terms of social network theory, this means that individuals manifestly recognized as enemies do not add to the structural tension of a community. The notion of gauge transformation is instrumental to understand this important feature of our social networks.

Computation of global balance A signed network is a graph in which the nodes are the users and the edges are their bipartite relationships. In this paper the relationships are always represented as mutual, i.e., the edges are undirected. For the 3 on-line social networks discussed in this paper (see Table 1 and the Material and Methods Section for details), this is largely an acceptable assumption. Practically, if $J_{ij} \in \{\pm 1\} = \mathbb{B}_2$ is the edge between the nodes s_i and s_j of the graph, computing global balance means assigning a $+1$ or a -1 to all the nodes so as to minimize the energy functional

$$h(\mathbf{s}) = \sum_{(i,j)} (1 - J_{ij} s_i s_j)/2, \qquad \textbf{[1]}$$

where the summation runs over all adjacent pairs of nodes and $\mathbf{s} = [s_1 \dots s_n]^T \in \mathbb{B}_2^n$ i.e., $s_i \in \{\pm 1\}$, $i = 1, \dots, n$, with $n =$ number of nodes. When J_{ij} represents friendship ($J_{ij} = +1$) each term in **[1]** gives a zero contribution if $\mathrm{sign}(s_i) = \mathrm{sign}(s_j)$ and a $+1$ contribution if $\mathrm{sign}(s_i) = -\mathrm{sign}(s_j)$, while when J_{ij} represents hostility ($J_{ij} = -1$) the summand is zero if $\mathrm{sign}(s_i) = -\mathrm{sign}(s_j)$ and $+1$ otherwise. The network is exactly balanced when there exists $\mathbf{s} \in \mathbb{B}_2^n$ such that all terms in **[1]** can be made simultaneously equal to zero. Call $\mathcal{J}$ the $n \times n$ matrix of entries J_{ij}. As the signed graph is undirected, its adjacency matrix $\mathcal{J}$ is symmetric: $J_{ij} = J_{ji}$. Up to a constant, **[1]** can be identified with the Hamiltonian of a so-called (heterogeneous) Edwards-Anderson spin glass, with bimodal bond distribution [38]. For connected signed graphs, if k_i is the connectivity degree of the i-th node, and $\mathcal{K} = \mathrm{diag}(k_1, \dots, k_n)$, then **[1]** is exactly balanced if and only if the smallest eigenvalue of the Laplacian $\mathcal{L} = \mathcal{K} - \mathcal{J}$ is equal to 0 [18]. If this is not the case, then necessarily no choice of $\mathbf{s}$ can render all terms in **[1]** simultaneously zero: $h(\mathbf{s}) > 0 \; \forall \mathbf{s} \in \mathbb{B}_2^n$. Computing the global balance of the network then means solving the following boolean optimization problem

$$\delta = \min_{\mathbf{s} \in \mathbb{B}_2^n} h(\mathbf{s}) = \min_{\mathbf{s} \in \mathbb{B}_2^n} \left(m - \frac{1}{2} \mathbf{s}^T \mathcal{J} \mathbf{s} \right), \qquad \textbf{[2]}$$

where m is the number of edges, $m = \sum_{i=1}^n k_i/2$. In correspondence of $\mathbf{s}_o = \mathrm{argmin}_{\mathbf{s} \in \mathbb{B}_2^n} h(\mathbf{s})$, the residual positive terms in **[1]** correspond to the least number of unbalanced pairwise relationships between nodes (i.e., the frustrations of the spin glass Hamiltonian **[1]** in its ground state).

The enormous dimension of the configuration space (2^n) makes the problem **[2]** hard to solve. As a matter of fact, MAX-CUT is one of those problems for which currently existing heuristics are normally tested only on small-medium benchmark problems, of the order of 10^3 nodes [31, 23, 24]. The novel heuristic we have introduced in [33] is however able to produce fairly tight upper and lower bounds for δ (henceforth δ_{up} and δ_{low}) also for very large signed graphs. This new local search algorithm is described in some detail in the Supplementary Notes and in [33]. The outcome of the algorithm is a gauge transformation of the adjacency matrix $\mathcal{J}$ into the equivalent $\mathcal{J}_\sigma$:

$$\mathcal{J} \to \mathcal{J}_\sigma = T_\sigma \mathcal{J} T_\sigma, \qquad \textbf{[3]}$$

where T_σ is a diagonal signature matrix $T_\sigma = \mathrm{diag}(\sigma)$, $\sigma \in \mathbb{B}_2^n$, such that $\mathcal{J}_\sigma$ has the same frustration as $\mathcal{J}$, but the least possible number of negative entries among all transformations of the form **[3]**. Since $\sigma = T_\sigma \mathbf{1}$, where $\mathbf{1}$ is the all spins up configuration, in terms of the energy function **[1]** we have

$$h(\sigma) = m - \frac{1}{2} \sigma^T \mathcal{J} \sigma = m - \frac{1}{2} \mathbf{1}^T T_\sigma \mathcal{J} T_\sigma \mathbf{1}, \qquad \textbf{[4]}$$

meaning that minimizing the energy over the spin configurations $\mathbf{s} \in \mathbb{B}_2^n$ as in **[2]** or minimizing the number of negative entries of $\mathcal{J}_\sigma$ through operations such as **[3]** yield identical results.

Global balance for social networks The local search algorithm was applied to the 3 on-line social networks of Table 1. Some ~ 4700 replicas were computed for Epinions, ~ 8000 for Slashdot and ~ 18000 for WikiElections. Of these replicas, the best (in terms of δ) 606 for Epinions, 953 for Slashdot, and 1000 for WikiElections were used in our statistics. The distributions of these $\delta_{up,replica}$ are shown in Fig. 2. The corresponding $\delta_{up} = \min(\delta_{up,replica})$ are given in Table 1, where also the corresponding lower bounds on δ, δ_{low}, are shown. That this algorithm scales well with size, and in fact that it can deal effectively with the signed social networks of dimension $\sim 10^5$ used in this paper, is proved by the tiny gap left between δ_{low} and δ_{up}, see Table 2, which guarantees that the estimate for δ is accurate. For 2 of the 3 networks, we have essentially computed the true optimum, as $\delta_{low}/\delta_{up} > 0.99$, while the residual gap in the third network (Slashdot, $\delta_{low}/\delta_{up} > 0.95$) is most likely due to the lack of precision of the lower bound computation (see Supplementary Notes for mode details).

By definition, a local optimum of the energy **[1]** is any $\mathbf{s}$ such that for every user the majority of pairwise relationships are "satisfied", i.e., yield a zero contribution to **[1]**. Due to the ruggedness of the energy landscape, the number of local minima can be huge [27]. It is only by solving **[2]** that a local minimum becomes also a global optimum and, in the present context, acquires the meaning of balance value for the network. Since our computed δ_{up} is very close to the true δ, essentially all the residual conflicts in $\mathcal{J}_\sigma$ are ineliminable, i.e., they represent the real disorder of the problem. Due to the gauge equivalence, what holds in the ground state $\mathbf{1}$ for $\mathcal{J}_\sigma$ holds also in the configuration $\mathbf{s}_o = \sigma$ for the original $\mathcal{J}$. In the optimal balance state $\mathbf{s}_o$, a consistent fraction of users results to be completely free from tensions: from the 52.7% of WikiElections to the 83.7% for Epinions, see Table 2. If we restrict to these users, then the identification of clusters of perfectly balanced subcommunities is straightforward, as it corresponds to determining the connected components of the subgraph of perfectly balanced users. See Supplementary Notes for details.

Statistical analysis of the level of balance For a signed graph, the amount of frustration depends on the topology of the network, on the percentage of negative edges and, most of all, on their distribution on the graph. Unlike for spin glasses on regular lattices, for heterogeneous signed networks systematic predictions of the expected frustration, given the connectivity and the percentage of negative edges, are completely missing. We observe that on the three social networks of Table 1 the fraction of negative edges is always limited ($q = m^-/m = 16.7 \div 23.6\%$). In terms of spin glasses, this would corresponds to a "partially ferromagnetic" quenching (more ferromagnetic than anti-ferromagnetic bonds). Obviously this leads to a lower frustration than in a spin glass with equally distributed edge signs. To evaluate if also the arrangements of the negative edges on the graph are favoring balance, we have to compare the sign arrangements on our networks with null models. In the null models we discuss here the edge signs are drawn as i.i.d. variables from a Bernoulli distribution with probability of negative sign equal to q, hereafter denoted $\mathcal{B}(q)$. For each of the 3 networks, 100 randomizations were performed, and the corresponding δ_{low}^{null} and δ_{up}^{null} computed solving **[2]** via the same heuristic used for the true networks (see Table 2). The distribution of the δ_{up}^{null} is compared with δ_{low} and δ_{up} of the true networks in Fig. 2. It can be observed that the null models are unavoidably much more frustrated than the real social networks (Z-test, p-value ~ 0). This means that indeed the organization of the signs in our social networks is such that tensions are largely avoided. Analogous results are obtained if the null models are constructed using a hypergeometric distribution, corresponding to reshuffling randomly the signs on the edges while keeping constant the ratio negative/positive edges, see Fig. S4.

For our networks, the property of being much more balanced than expected goes beyond the statistical significance of a Z-test on null models. As a matter of fact, δ_{up} is even less than a Shannon-type bound which can be associated with the average frustration of our null models. For n and m sufficiently large, denoting $R = n/m$ the rate and $D = \delta/m$ the distortion, the rate-distortion theorem (see [39]) affirms that when the edge signs are drawn as i.i.d. variables from $\mathcal{B}(q)$ then the distortions achievable are in expectation lower bounded by the distortion-rate curves shown in Fig. 3, regardless of the topology of $\mathcal{J}$, see Supplementary Notes for a more rigorous formulation of these information-theoretical concepts. Distortions (and hence frustrations) that lie below this Shannon bound must be considered as obtained from edge sign assignments that are highly atypical for the probability "source" $\mathcal{B}(q)$. All three networks have sign arrangements that violate the Shannon bound, meaning that indeed the true "quenchings" are away from $\mathcal{B}(q)$ with high significance. In Fig. 3 notice that, instead, the distortions δ_{low}^{null}/m and δ_{up}^{null}/m of the null models all lie above the Shannon bounds, as expected.

Skewness of the sign distributions and its social meaning The feature that makes our networks so atypical is the skewness of the sign distribution on the individuals. In particular, the three networks have a significant fraction of nodes that are enriched for positive or negative edges (cumulative binomial test, p-value 10^{-5}), property not shared with the null models, see Fig. 4 and Table S5. Both fat tails of this sign distribution contribute to increase the balance of a network: the tail of positive edges because users with many friends have less enemies than expected from null models; the tail of negative edges for the opposite reason. A direct consequence of the sign skewness is that a considerable part of negative edges can be eliminated by means of gauge transformations, meaning that a vast fraction of the negative edges contribute only to the apparent disorder, not to the real frustration. On the contrary, the reduction of negative edges in the null models is always minimal, see Fig. 2 and Table 2.

That the reduction of negative edges in passing from $\mathcal{J}$ to $\mathcal{J}_\sigma$ is primarily due to users with high connectivity of enemies is confirmed on all 3 networks by the signed degree distributions of Fig. S1-S3 (compare the degree distributions of negative edges in $\mathcal{J}$ and $\mathcal{J}_\sigma$). In practice, a small fraction of individuals attracting a large number of negative edges contributes less to unbalance the social community than a homogeneous distribution of unfriendly relationships. The sociological interpretation of this fact is clear: unpopular individuals are easily "cast away" from the bulk of the community without creating much conflict within the community itself. Something similar does not happen for homogeneous distributions of the negative edges in the community. In conclusion, in all three networks analyzed, the local process of choosing friends/enemies induces a collective behavior which is strongly biased towards the creation of a disorder that is only apparent, thereby confirming the validity of Heider's theory for this class of networks.

Materials and Methods

The 3 signed social networks analyzed in this study were downloaded from the Stanford Network Analysis Platform (http://snap.stanford.edu/) [12]:

- *Epinions*: trust/distrust network among users of product review web site Epinions [40], [13];
- *Slashdot*: friend/foes network of the technological news site Slashdot (Zoo feature) [41], [12];
- *WikiElections*: election of admin among Wikipedia users [42].

More details on these networks are provided in [13], see also [12] for Slashdot. The size (n) and number of edges (m) of these networks are given in Table 1. The edges of the networks are always considered as undirected. This process leads to only a limited number of sign inconsistencies between pairs of edges J_{ij} and J_{ij}, see Table S1. These inconsistencies are disregarded in our analysis. The methods used in the paper are described in full detail on the Supplementary Notes.

1. **Szell, M, Lambiotte, R, & Thurner, S. (2010) Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences* 107, 13636–13641.**
2. **Newman, M. E. J, Barabási, A. L, & Watts, D. J, eds. (2006) *The Structure and Dynamics of Networks*. (Princeton University Press).**
3. **Hogg, T, Wilkinson, D, Szabo, G, & Brzozowski, M. (2008) *Multiple relationship types in online communities and social networks*, AAAI SIP 2008.**
4. **Palla, G, Barabasi, A.-L, & Vicsek, T. (2007) Quantifying social group evolution. *Nature* 446, 664.**
5. **Albert, R & Barabasi, A.-L. (2000) Topology of evolving networks: local events and universality. *Phys.Rev.Lett.* 85, 5234.**
6. **Milo, R, Shen-Orr, S, Itzkovitz, S, Kashtan, N, Chklovskii, D, & Alon, U. (2002) Network motifs: Simple building blocks of complex networks. *Science* 298, 824–827.**
7. **Vega-Redondo, F. (2007) *Complex social networks*, Econometric Society monographs. (Cambridge Univ. Press, Cambridge).**
8. **Newman, M. E. J. (2006) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103, 8577–8582.**
9. **Borgatti, S. P, Mehra, A, Brass, D. J, & Labianca, G. (2009) Network analysis in the social sciences. *Science* 323, 892–895.**
10. **Wasserman, S & Faust, K. (1994) *Social Network Analysis: methods and applications*. (Cambridge Univ. Press).**
11. **Easley, D & Kleinberg, J. (2010) *Networks, Crowds, and Markets. Reasoning About a Highly Connected World*. (Cambridge Univ. Press, Cambridge).**
12. **Kunegis, J, Lommatzsch, A, & Bauckhage, C. (2009) *The Slashdot Zoo: Mining a Social Network with Negative Edges*. pp. 741–741.**
13. **Leskovec, J, Huttenlocher, D, & Kleinberg, J. (2010) *Signed Networks in Social Media*. (Atlanta, GA).**
14. **Srinivasan, A. (2011) Local balancing influences global structure in social networks. *Proceedings of the National Academy of Sciences* 108, 1751–1752.**

15. Heider, F. (1946) Attitudes and cognitive organization. *Journal of Psychology* 21, 107–122.
16. Cartwright, D & Harrary, F. (1956) Structural balance: a generalization of Heider's theory. *Psychological Review* 63, 277–292.
17. Harary, F & Kabell, J. A. (1980) A simple algorithm to detect balance in signed graphs. *Mathematical Social Sciences* 1, 131 – 136.
18. Kunegis, J, Schmidt, S, Lommatzsch, A, Lerner, J, W., D. L. E, & Albayrak, S. (2010) *Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization.* pp. 559–559.
19. Maybee, J. S & Maybee, S. J. (1983) An algorithm for identifying Morishima and anti-Morishima matrices and balanced digraphs. *Mathematical Social Sciences* 6, 99 – 103.
20. Harary, F. (1959) On the measurement of structural balance. *Behavioral Science* 4, 316–323.
21. Harary, F. (1960) A matrix criterion for structural balance. *Naval Res. Logist. Quart.* 7, 195–199.
22. Barahona, F. (1982) On the computational complexity of Ising spin glass models. *J. Phys A: Math. Gen* 15, 3241–3253.
23. Commander, C. W. (2009) in *Encyclopedia of Optimization*, eds. Floudas, C. A & Pardalos, P. M. (Springer US), pp. 1991–1999.
24. Festa, P, Pardalos, P, Resende, M, & Ribeiro, C. (2002) Randomized heuristics for the max-cut problem. *Optimization Methods and Software* 17, 1033–1058.
25. Mezard, M & Montanari, A. (2009) *Information, Physics, and Computation*. (Oxford University Press, Inc., New York, NY, USA).
26. Galam, S. (1996) Fragmentation versus stability in bimodal coalitions. *Physica A: Statistical and Theoretical Physics* 230, 174 – 188.
27. Antal, T, Krapivsky, P. L, & Redner, S. (2005) Dynamics of social balance on networks. *Phys. Rev. E* 72, 036121.
28. Marvel, S. A, Strogatz, S. H, & Kleinberg, J. M. (2009) Energy landscape of social balance. *Phys. Rev. Lett.* 103, 198701.
29. Hartmann, A. K & Rieger, H. (2001) *Optimization Algorithms in Physics*. (Wiley-VCH).
30. Martin, O. C. (2005) in *New Optimization Algorithms in Physics*, eds. Hartmann, A. K & Rieger, H. (Wiley-VCH).
31. Burer, S, Monteiro, R, & Zhang, Y. (2002) Rank-two relaxation heuristics for max-cut and other binary quadratic programs. *SIAM Journal on Optimization* 12, 503–521.
32. Palagi, L, Piccialli, V, Rendl, F, Rinaldi, G, & Wiegele, A. (2010) in *Handbook on Semidefinite, Cone and Polynomial Optimization: Theory, Algorithms, Software and Applications.*, eds. Anjos, M. F & Lasserre, J. B.
33. Iacono, G, Ramezani, F, Soranzo, N, & Altafini, C. (2010) Determining the distance to monotonicity of a biological network: a graph-theoretical approach. *IET Systems Biology* 4, 223–235.
34. Toulouse, G. (1977) Theory of the frustration effect in spin glasses : I. *Communications on Physics* 2, 115.
35. Zaslavsky, T. (1982) Signed graphs. *Discrete Appl. Math.* 4, 47–74.
36. Ciliberti, S & Mézard, M. (2005) The theoretical capacity of the parity source coder. *J. Stat. Mech.* p. P10003.
37. Wainwright, M. J, Maneva, E, & Martinian, E. (2010) Lossy Source Compression Using Low-Density Generator Matrix Codes: Analysis and Algorithms. *IEEE Tr. Inf. Th.* 56, 1351–1368.
38. Binder, K & Young, A. P. (1986) Spin glasses: Experimental facts, theoretical concepts, and open questions. *Rev. Mod. Phys.* 58, 801–976.
39. Cover, T. M & Thomas, J. A. (2006) *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. (Wiley-Interscience).
40. Guha, R, Kumar, R, Raghavan, P, & Tomkins, A. (2004) *Propagation of trust and distrust*, WWW '04. pp. 403–412.
41. Lampe, C. A, Johnston, E, & Resnick, P. (2007) *Follow the reader: filtering comments on Slashdot*, CHI '07. pp. 1253–1262.
42. Burke, M & Kraut, R. (2008) *Mopping up: modeling Wikipedia promotion decisions*, CSCW '08. pp. 27–36.
43. Goemans, M & Williamson, D. (1995) Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM* 42, 1115–1145.

Table 1. Signed social networks

Network	n	m	m^-	m^+	q	R
Epinions	131513	708507	118619	589888	0.167	0.186
Slashdot	82062	498532	117599	380933	0.236	0.165
WikiElections	7114	100321	21529	78792	0.214	0.071

Data for the 3 networks described in Materials and Methods, after symmetrization (the original directed graphs are reported in Table S1). n and m are the number of nodes and edges of the undirected graph, m^- and m^+ are the n. of negative and positive edges of the networks. $q = m^-/m$ is the probability of a negative edge and $R = n/m$ is the "rate of compression" (see text and Supplementary Notes).

Table 2. Global balance of the networks

Network	δ_{low}	δ_{up}	δ_{low}^{null}	δ_{up}^{null}	δ_{up}/m	δ_{low}/δ_{up}	perfectly balanced nodes
Epinions	50452	50806	105247	105520	0.0717	0.9930	110087 (83.71%)
Slashdot	70014	73604	90346	106163	0.1476	0.9512	56041 (68.29%)
WikiElections	14194	14245	20878	20880	0.1420	0.9964	3766 (52.94 %)

δ_{low} and δ_{up} are the lower and upper bounds on the global balance. These are much lower than δ_{low}^{null} and δ_{up}^{null}, the corresponding average values of balance obtained on null models generated from a Bernoulli distribution $\mathcal{B}(q)$. The ratio δ/m (the more conservative δ_{up}/m for us) represents the distortion, i.e., the fraction of frustrated bipartite relationships in the global balance configuration $\mathbf{s}_o$, see Fig. 3. For the values of δ_{low} and δ_{up}, the ratio δ_{low}/δ_{up} is much higher than the value achieved by popular semidefinite programming approaches to MAX-CUT (0.8785, see [43]), meaning that our ground state algorithm is indeed quite efficient. The last column reports the number and the percentage of perfectly balanced nodes in the ground state.

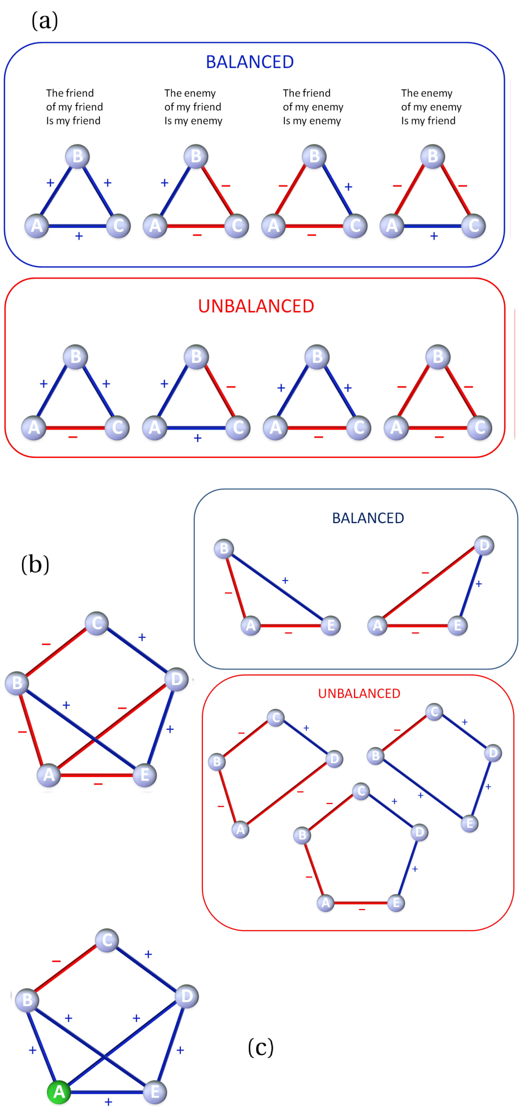
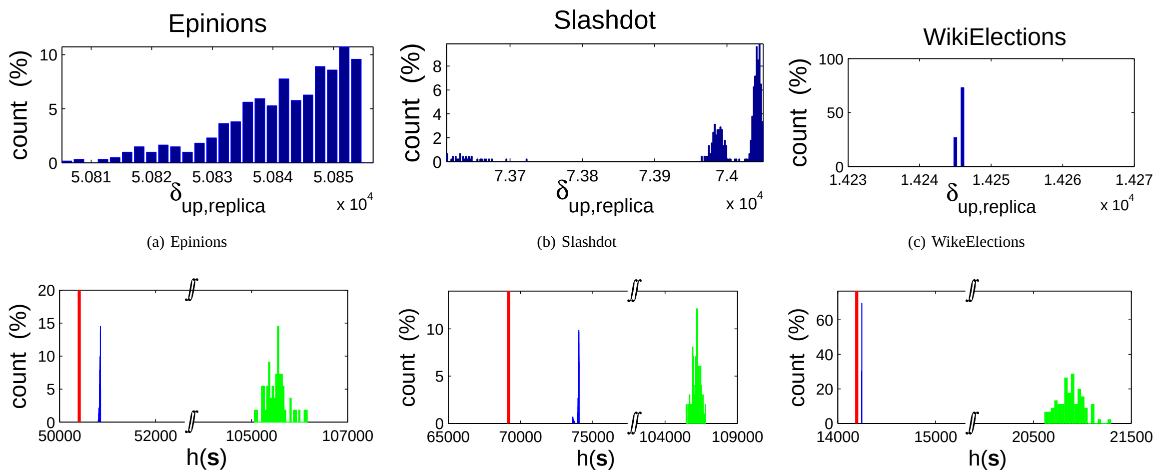


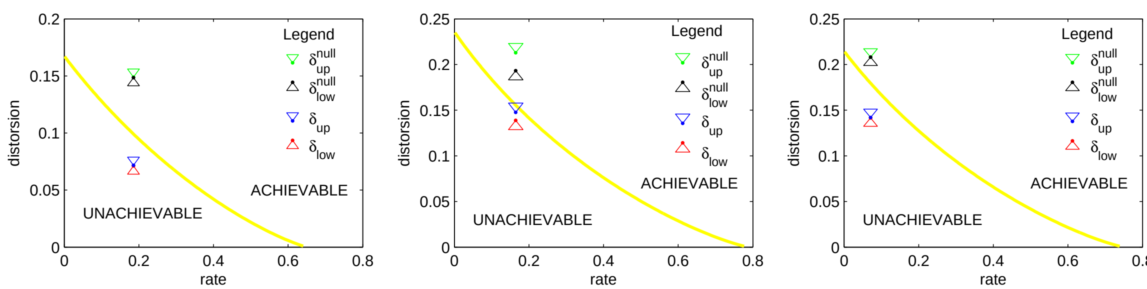
Fig. 1. Balanced and unbalanced graphs. (a): Simplest cases of balance and unbalance: triangles. Users A and C are related directly and indirectly (through B). The sentences on the top connote this indirect relationship between A and C. Blue edges represent friendship, red hostility. The triangles are balanced when the direct and the indirect relationships have the same sign, unbalanced otherwise. (b): For generic graphs, testing all triangles may not give a satisfactory measure of the global balance. In the example the graph is not globally balanced, although all triangles are balanced. (c): Illustration of a gauge transformation. Applying a sign change to all edges adjacent to the node A of (b) only one negative edge is left in the graph, and $\delta = 1$ in this case. In **[2]** this optimum corresponds to choosing $s_A = -1, s_B = s_C = s_D = s_E = +1$. Notice that counting the total number of balanced/unbalanced cycles is not a significant measure of balance. It is evident, then, that in the ground state the nodes A, D, and E are perfectly balanced while the nodes B and C have a nonnegative sum of signed edges (i.e., the global optimum is also a local optimum for each node.).



(a) Epinions (b) Slashdot (c) WikeElections

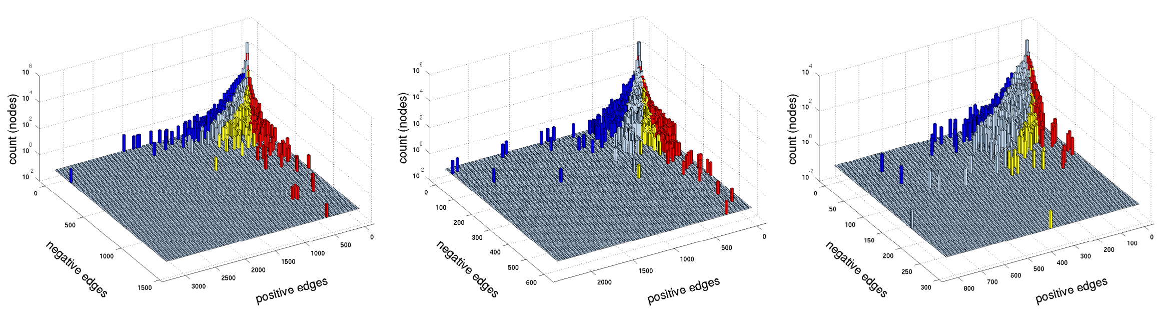
(d) Epinions (e) Slashdot (f) WikeElections

Fig. 2. Global balance and its statistical significance. Top row: optimal level of balance $\delta_{up,replica}$ reached on different replicas for the 3 networks. The (low energy) replicas shown are 606 for Epinions, 953 for Slashdot, and 1000 for WikiElections. Bottom row: comparison of $\delta_{up,replica}$ with lower bounds and with null models generated from a Bernoulli distribution $\mathcal{B}(q)$. The lower bounds δ_{low} are shown in red, the distributions of $\delta_{up,replica}$ are in blue, and the distribution of δ_{up}^{null} in 100 null models in green. In each of the 3 networks $\delta_{up} = \min(\delta_{up,replica}) \ll \delta_{up}^{null}$, meaning that the true networks are much less frustrated than expected from the null models (Z-test, with p-value $< 10^{-100}$). Furthermore, the interval of uncertainty of the optimal level of balance is very limited, since $\delta_{low}/\delta_{up} > 0.95$ ($\delta_{low}/\delta_{up} > 0.99$ for Epinions and WikiElections) and $\delta_{up} - \delta_{low} \ll \delta_{up}^{null} - \delta_{up}$.



(a) Epinions (b) Slashdot (c) WikiElections

Fig. 3. Rate-distortion plots. In the rate ($R = n/m$) - distortion ($D = \delta/m$) plane, the yellow curves are the Shannon bounds of the rate-distortion theorem associated to a Bernoulli distribution $\mathcal{B}(q)$. The region above (below) the curve is achievable (unachievable) in expectation by an edge sign assignment drawn as a length-m sequence of i.i.d. variables from $\mathcal{B}(q)$, compressed to a length-n sequence, and then reconstructed. The compression step is equivalent to our ground state search problem, and the distortion obtained is the frustration normalized by m. Full details on these information-theoretical aspects are provided in the Supplementary Notes. The distortion of the three true sign assignments (tip of the triangles, blue for δ_{up} and red for δ_{low}, partially overlapping) is less than this Shannon-type bound, meaning that these edge signatures are significantly away from a typical i.i.d. sequence from $\mathcal{B}(q)$. The signatures used in the null models of Table 2 (tip of the triangles, green for δ_{up}^{null} and black for δ_{low}^{null}, also partially overlapping) are instead in the achievable region.



(a) Epinions (b) Slashdot (c) WikiElections

Fig. 4. Global balance and sign skewness. All three networks have a significant percentage of nodes enriched for positive (blue) or negative edges (yellow and red), see Table S5. The sign skewness of a node is computed through a cumulative binomial test (p-value 10^{-5}, see Supplementary Notes). In particular, the nodes in red are adjacent to more negative than positive edges in $\mathcal{J}$. Gauge transforming these nodes reduced considerably the amount of negative edges of the networks while not altering their frustration. These histograms should be compared with the corresponding histogram for a null model, shown in Fig. S7.