

Exploring the low-energy landscape of large-scale signed social networks

G. Facchetti, G. Iacono, C. Altafini*
SISSA, via Bonomea 265, 34136 Trieste, Italy.

Analogously to a spin glass, a large-scale signed social network is characterized by the presence of disorder, expressed in this context (and in the social network literature) by the concept of structural balance. If, as we have recently shown, the signed social networks currently available have a limited amount of true disorder (or frustration), it is also interesting to investigate how this frustration is organized, by exploring the landscape of near-optimal structural balance. What we obtain in this paper is that while one of the networks analyzed shows a unique valley of minima, and a funneled landscape that gradually and smoothly worsens as we move away from the optimum, another network shows instead several distinct valleys of optimal or near-optimal structural balance, separated by energy barriers determined by internally balanced sub-communities of users, a phenomenon similar to the replica-symmetry breaking of spin glasses. Multiple, essentially isoenergetic, arrangements of these communities are possible. Passing from one valley to another requires to destroy the internal arrangement of these balanced subcommunities and then to reform it again. It is essentially this process of breaking the internal balance of the subcommunities which gives rise to the energy barriers.

PACS numbers: 89.20.-a, 89.20.Hh, 89.75.Fb, 75.10.Nr

I. INTRODUCTION

Signed social networks represent an interesting class of networks at the interface between network theory, social sciences, and statistical physics [1–7]. In these networks, the nodes represent the individuals and the edges can assume a positive or a negative sign according to the type of relationship established between pairs of individuals: a positive edge represents a form of friendship/collaboration/trust (“I like”), while a negative edge represents aversion/competition/mistrust (“Don’t like”).

Like in an Ising spin glass, the presence of negative edges introduces *frustration* in the network, which for social networks is associated to a property called *structural balance*, introduced by F. Heider in [8], then generalized and formulated in graphical terms by Cartwright-Harary in [9]. According to the theory proposed by Heider, individuals tend to establish relations that avoid tensions: this can be condensed in the statement “*the enemy of my enemy is my friend*” (and similar, equivalent statements, see e.g. Fig. 1 of [2]). When these social interactions are described through signed networks, the theory can be reformulated in terms of frustration of the cycles which are present in the corresponding signed graphs [9]. In particular, all cycles with an odd number of negative edges (hereafter called negative cycles) are frustrated: they do not obey to Heider’s principle and they contribute to the global frustration of the whole network. Since the concept of frustration is formally identical to the one used in spin glass theory [10], also

the notion of structural balance admits a statistical physics analog. Exact structural balance, in particular, corresponds to lack of frustration, as in a Mattis model [11]. More generally, computing the structural balance of a signed social network corresponds to computing the ground state of a (non-planar) Ising spin glass, which is a well-known NP-hard problem [12, 13]. In fact, in social network theory a common approach to overcome this difficulty has been to focus on the simplest cyclic motif potentially encoding frustration, namely triangles of pairwise relations, comparing the frequency of positive and negative triangles [7, 14]. However, the determination of the level of balance cannot always be reduced to the count and characterization of local motifs. If for all-to-all networks [1, 5] triangles are a reliable measure of structural balance, for heterogeneous networks (like real social networks) it is not clear to what extent counting triangles reflects the structural balance (which is a *global* property [6], in the Cartwright-Harary generalization).

For this purpose, in [2], exploiting the analogy between signed social networks and Ising spin glasses we have computed structural balance using algorithms inspired by the literature on ground state search. On the signed social networks currently available (which can reach sizes of 10^5 nodes), these calculations have proved quite effective (with a guaranteed precision of 5% in the value of structural balance), and have allowed to conclude that these networks are extremely balanced, much more than expected from random edge sign distributions. In spin glass terminology, this result can be restated by saying that the social networks have a limited amount of *true disorder* (and a significant amount of *apparent disorder*).

The algorithm we have used (described in some

*Corresponding author: altafini@sissa.it

detail in [2], see also [15]) has an heuristic character, hence in order to obtain reliable results it has to be run many times, changing the initial conditions and the random seeds. In statistical physics terminology, each such run corresponds to a *replica* of the system. As the “cooling” schedule differs from replica to replica, the level of structural balance, i.e., the energy reached (which typically corresponds to a local minimum) may differ from one replica to the other.

The purpose of this paper is to use the large amount of independent replicas produced in this way to try to obtain a description of the low-energy landscape of the system i.e., of the landscape of near optimal balance of the social networks. In statistical physics, in presence of frustration the low-energy landscape is usually rugged and contains a lot of useful information on the properties of the spin glass. For example if the system has multiple distinct valleys of nearly identical energy separated by energetic barriers, then it is said to have a replica symmetry breaking picture [16], meaning that ergodicity is broken and different cooling procedures lead to different, isoenergetic (or nearly isoenergetic), minima separated by energy barriers. Some other times, instead, it may happen that the landscape is funneled around a single “dominant” valley, possibly surrounded by a plethora of local minima of little significance (and progressively higher energy). The two different pictures are both observed on the two social networks considered in this study. Most interestingly, we can also obtain insight on the origin of these differences in the landscape. In the case of multiple competing valleys, in fact, it is possible to isolate internally balanced subcommunities whose orientation with respect to the rest of the network changes passing from one valley to another. The sample paths connecting one valley to the other require to destroy the internal balance of (some of) these subcommunities and then to reform it again in the opposite orientation. It appears that the energy barriers between the near-optimal valleys is largely due to this process of breaking the internal balance of the subcommunities. In the case of funneled landscape, no internally balanced subcommunity can instead be observed. In any case, in both pictures all low-energy replicas are sufficiently close to each other, situation which corresponds to a “partially ferromagnetic” spin glass, coherently with the large fraction of positive edges, as well as with the low frustration that characterizes these networks.

TABLE 1: Social networks. Main features of the two networks: n and m are the number of nodes and edges of the undirected graph. The 2-component subnetwork has been obtained considering only the largest connected component and removing all leaves. In this paper relationships are always represented as mutual, i.e., the edges are undirected.

original network					
Network	n	m	n. of –	n. of +	
<i>Epinions</i>	131828	841372	123705	717667	
<i>Slashdot</i>	82144	549202	124130	425072	
largest 2-component subnetwork					
Network	n	m	n. of –	n. of +	%–
<i>Epinions</i>	59235	641734	106286	535448	16.5
<i>Slashdot</i>	52048	468523	112023	356500	23.9

II. THE SIGNED SOCIAL NETWORKS AND THEIR STRUCTURAL BALANCE

In this work we consider the following two signed social networks

- *Epinions*: trust/distrust network among users of product review web site Epinions [14, 17];
- *Slashdot*: friend/foes network of the technological news site Slashdot (Zoo feature) [18, 19],

both downloadable from the Stanford Network Analysis Platform (<http://snap.stanford.edu/>). Numbers of nodes (n), edges (m) and edge signs are provided in Table 1, where we report also the data for the largest 2-component subnetwork on which all the analysis has been performed. Indeed, leaves can be removed since they are not involved in cycles, hence they do not contribute to frustration. Further details on these networks are provided in [14], see also [19] for Slashdot.

Call $\mathbf{s} = [s_1 \dots s_n]^T$ with $s_i \in \{\pm 1\} = \mathbb{B}_2$, $i = 1, \dots, n$ the “spin” variables associated to the nodes (individuals) of the network. Let also \mathcal{J} be the $n \times n$ symmetric matrix of entries J_{ij} which represent the undirected relationships between nodes s_i and s_j (friendship: $J_{ij} = +1$; hostility: $J_{ij} = -1$). Computing the global structural balance means assigning a $+1$ or a -1 to the nodes in such a way to minimize the energy function:

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

where the summation runs over all adjacent pairs of nodes. In matrix form (1) is

$$h(\mathbf{s}) = m - \frac{1}{2} \mathbf{s}^T \mathcal{J} \mathbf{s}.$$

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. If this is not the case, then global balance becomes the solution of a Boolean optimization problem

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

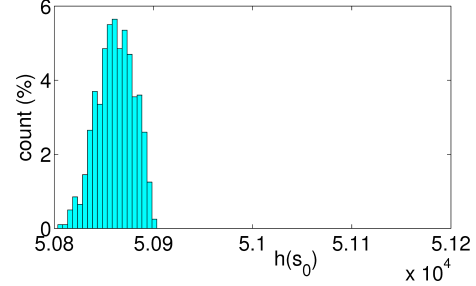
In correspondence of $\mathbf{s}_o = \operatorname{argmin}_{\mathbf{s} \in \mathbb{B}_2^n} h(\mathbf{s})$, the residual positive terms in (1) correspond to the least number of unbalanced pairwise relationships between nodes (i.e., the frustrations of the spin glass Hamiltonian (1) in its ground state).

TABLE 2: Structural balance. For the networks of Table 1, we report a lower and an upper bound (δ_{low} and δ_{up}) on the true structural balance (data from [2]). The high ratio $\delta_{\text{up}}/\delta_{\text{low}}$ shows that indeed our calculations are approaching sufficiently close the true ground state energy. By construction, $\delta_{\text{up}} = \min_k \delta^{(k)}$ over all replicas considered, and only the r replicas of energy $\delta^{(k)} \in [\delta_{\text{up}}, \delta_{\text{up}} + \epsilon]$ are considered.

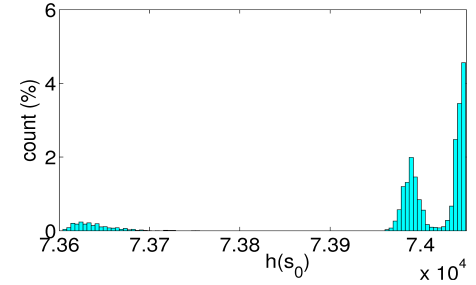
	δ_{low}	δ_{up}	$\delta_{\text{up}}/\delta_{\text{low}}$	ϵ	r
<i>Epinions</i>	50452	50806	0.9922	310	606
<i>Slashdot</i>	70014	73604	0.9512	450	5557

For the networks considered in this study an heuristic estimate of δ was carried out in [2]. Recall from [2] that the algorithm we apply uses the idea of gauge transformations as a way to progressively eliminate the apparent disorder in the adjacency matrix \mathcal{J} , see [2] for the details. The outcome of the algorithm is a gauge-equivalent adjacency matrix $\mathcal{J}_\sigma = T_\sigma \mathcal{J} T_\sigma$, where the diagonal matrix $T_\sigma = \operatorname{diag}(\sigma)$, $\sigma = [\sigma_1, \dots, \sigma_n]$, $\sigma_i = \pm 1$, is chosen so as to minimize the number of negative entries of \mathcal{J}_σ (transformations $\mathcal{J} \rightarrow \mathcal{J}_\sigma$ are called gauge transformations in the statistical physics literature [20]). The practical (iterative) heuristic construction of T_σ is explained in the Supplementary Notes of [2].

The optimization procedure is repeated many times, starting from different initial points (independently and randomly chosen). Denote $\delta_{\text{up}} = \min_k \delta^{(k)}$ the best estimate obtained for the ground state energy over all replicas. Consider an energy band of width ϵ above δ_{up} . For both networks $\epsilon/\delta_{\text{up}} \sim 0.6\%$ is chosen, meaning a few hundreds of energy levels above δ_{up} are considered, see Table 2. Only replicas reaching the interval $[\delta_{\text{up}}, \delta_{\text{up}} + \epsilon]$



(a) Epinions



(b) Slashdot

FIG. 1: Distribution of the energies for all near-optimal replicas. The leftmost bar of each histogram corresponds to δ_{up} .

are retained for further analysis. The number of such replicas is $r = 606$ for Epinions and $r = 5557$ for Slashdot, see Table 2. The tight gap between the lower (δ_{low}) and upper (δ_{up}) estimate for δ (see Table 2), guarantees that these replicas are indeed low-energy minima sufficiently close to the true value of structural balance.

The histogram of the values of energies and the (relative) Hamming distances between these minima are reported respectively in Fig. 1 and in Fig. 2 for the two networks. A first comparison of these energies shows a significant difference between the two networks: the histogram for Epinions reports only a single broad group whereas the histogram for Slashdot shows three peaks, see Fig. 1. If we compute the relative Hamming distances between each pair of replicas (see Fig. 2), the two networks still present a different behavior: for Epinions the minima are all close one to the other (Hamming distances are distributed like a single Gaussian peak); on the contrary, several peaks can be identified among the minima we have found for Slashdot.

III. COMPUTATION OF THE LOW-ENERGY SPIN PATTERNS

Since we can expect that each replica represents a different local minimum in the energy landscape, we can explore the distribution of the frequencies of the sign with which a spin appears in the r replicas:

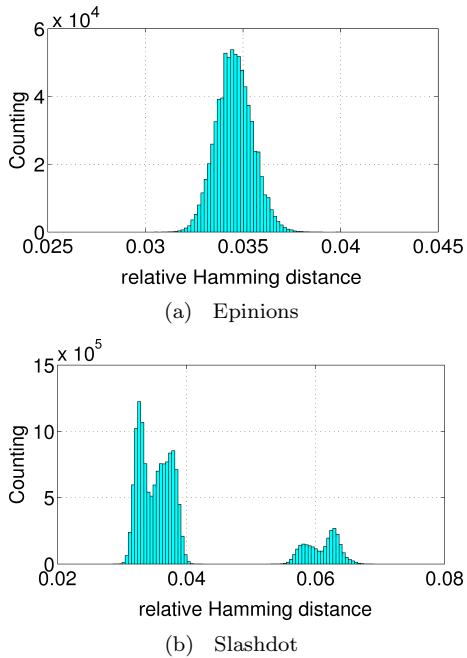


FIG. 2: Relative Hamming distance between low-energy replicas.

in particular, if $\mathbf{s}^{(1)}, \dots, \mathbf{s}^{(r)}$ are the configuration states of the minima in the r replicas, for each spin i we can define

$$\nu_i = \frac{1}{r} \min \left\{ \sum_{k=1}^r \frac{s_i^{(k)} - 1}{2}, \sum_{k=1}^r \frac{s_i^{(k)} + 1}{2} \right\},$$

where the two ratios represent the number of replicas in which the i -th spin has negative and positive sign respectively. For each value of ν between 0 and 1/2 (the histograms stop at $\nu = 1/2$ because of the global spin flip symmetry), the subset $W(\nu)$ of spins which have $\nu_i = \nu$ can be identified. For the subgraph corresponding to $W(\nu)$, we calculate: (i) the number of spins $n(\nu)$; (ii) the number of connected components $c(\nu)$; (iii) the maximal size of these connected components $z(\nu)$.

Exploring the distribution of the frequency index $c(\nu)$ (see the second panel in each row of Fig. 3), we can observe that both networks have broad peaks around frequencies $\nu = 1/3$ and $\nu = 1/2$, meaning that the corresponding spins appear flipped in a half or in a third of the low-energy replicas. The peak at $\nu = 0$ corresponds to spins with equal orientation in all replicas, hence it constitutes a fixed “backbone” which does not contribute to the variability.

A characterization of the two peaks at $\nu = 1/2$ and $\nu = 1/3$ in the plot of the index $c(\nu)$ can be carried out through a probabilistic model. By construction, the replicas are statistically independent. We may also expect that the flipping com-

TABLE 3: Probabilistic binomial model for the spin frequencies. Parameters of the fitting of the binomial probability distribution to the broad peaks of $c(\nu)$ of spin frequencies $\nu = 1/2$ and $\nu = 1/3$, see Fig. 3 (second panel of each row). Ranges are expressed as relative to the total number of replicas. For Epinions the selected intervals correspond respectively to $\nu \in [0.46, 0.50]$ and $\nu \in [0.31, 0.36]$. For Slashdot, the test yields the intervals $\nu \in [0.49, 0.50]$ and $\nu \in [0.32, 0.34]$.

<i>Epinions</i>		
<i>parameter</i>	$\nu = 1/2$	$\nu = 1/3$
<i>k/r</i> range for fitting	$[\frac{263}{606}, \frac{303}{606}]$	$[\frac{172}{606}, \frac{232}{606}]$
fitting parameter q	2378	642
r.m.s.d.	8	4
α	0.05	0.10
selected region	$[\frac{279}{606}, \frac{303}{606}]$	$[\frac{187}{606}, \frac{217}{606}]$
n. selected components	2336	528
<i>Slashdot</i>		
<i>parameter</i>	$\nu = 1/2$	$\nu = 1/3$
<i>k/r</i> range for fitting	$[\frac{2648}{5557}, \frac{2778}{5557}]$	$[\frac{1762}{5557}, \frac{1942}{5557}]$
fitting parameter q	1740	447
r.m.s.d.	3	1
α	0.10	0.10
selected region	$[\frac{2717}{5557}, \frac{2778}{5557}]$	$[\frac{1807}{5557}, \frac{1897}{5557}]$
n. selected components	1656	419

ponents which belong to the peak at $\nu = 1/2$ are independent with respect to the flipping components around $\nu = 1/3$ (and viceversa). Under these assumptions, the two peaks can be modeled separately (the plot of $c(\nu)$ is then view as the overlap of two curves), considering the Binomial probability distribution $B(r, p)$ as the theoretical distribution for the variable $r\nu$ for each peak:

$$\mathbb{P}[r\nu = k] = \binom{r}{k} p^k (1-p)^{r-k} \quad (3)$$

(where the value of p is fixed at $p = 1/2$ and $p = 1/3$ for the two peaks). If $Y(k)$ is a random variable representing the number of connected components (over a population of q flipping components, q unknown) having $\nu = k/r$, and assuming that each connected component can flip independently, then

$$Y(k) \sim B(q, \mathbb{P}[r\nu = k]).$$

By (3), its expectation value is

$$\mathbb{E}[Y(k)] = q \binom{r}{k} p^k (1-p)^{r-k},$$

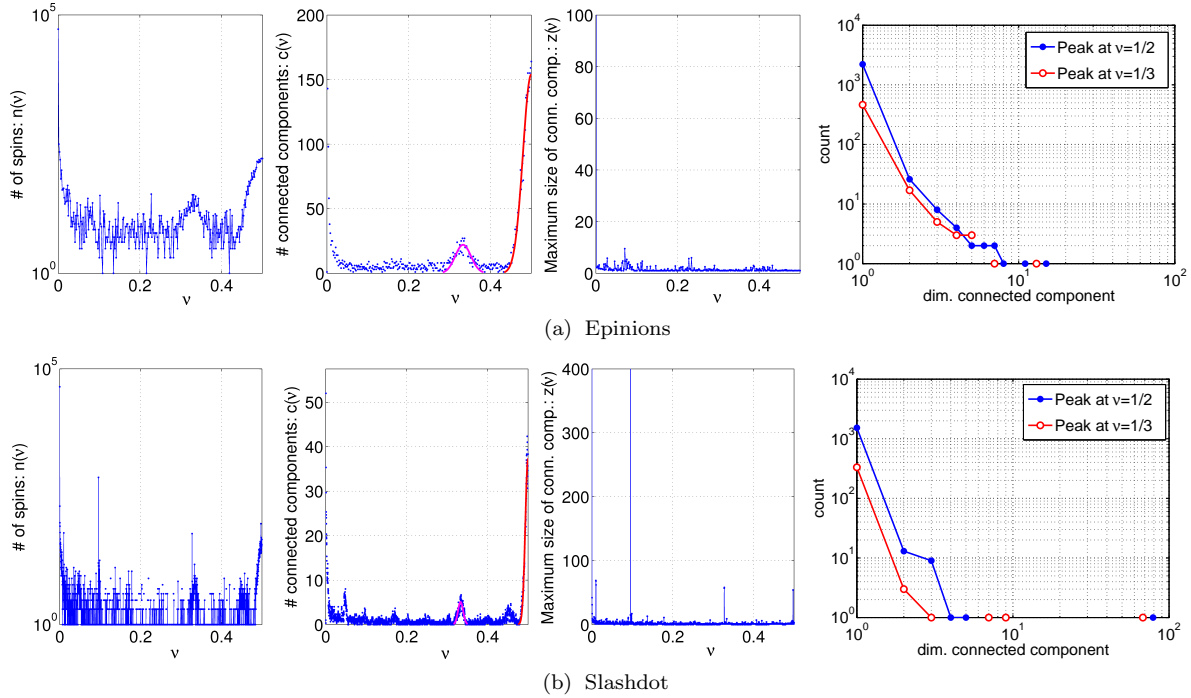


FIG. 3: Indices $n(\nu)$, $c(\nu)$ and $z(\nu)$ for the two networks. The rightmost panel in each row reports the dimensions of the connected components identified selecting the spins under the two peaks of $c(\nu)$ centered at $\nu = 1/3$ and at $\nu = 1/2$.

where the parameter q can be obtained by linear regression, fitting the experimental data with k varying inside the support of the peak. Once a level of significance α has been fixed, we can select only spins which are in the central part of the fitted peak with p-value less than α , as determined through an hypothesis test. The selected values of ν include only regions where the fitting error is smaller or at least comparable with the root mean square deviation (r.m.s.d.) of the data at each ν .

Extrapolating the connected subgraphs under the two peaks (Table 3 reports the results of the probabilistic model applied to the interpolation of these peaks), it is possible to obtain a statistics of the recurrent motifs which form most of the variability of the spin configurations around a ground state. A catalog of isoenergetic motifs (isoenergetic alternatives for a ground state) we identified through this interpolation is provided in Tables 6-7 (see also the Appendix for a more detailed characterization of these flipping motifs).

IV. ANALYSIS OF THE LOW-ENERGY LANDSCAPES FOR THE TWO NETWORKS

A. Epinions: a single valley.

We can deduce from the histograms of Fig. 1 and 2 that all the minima of the Epinions network belong to the same valley. As we move away from the bottom of the valley (global minimum), the energy tends to grow. This feature can be deduced from the analysis of the distance among replicas versus energy of a replica reported in Fig. 4(a). The 2D bar plot shows that, moving away from the minimum of energy, the Hamming distance increases monotonically. This means that the valley is characterized by sufficiently regular ascending walls, and that the basin of attraction of the global minimum is rather broad.

The high degeneracy of the ground state, which is also suggested by the sample trajectories of Fig. 4(b), is a well-known feature of Ising spin glasses with bimodal bonds [21]. The flipping of the isoenergetic motifs we have identified (see Tables 6-7 and Appendix), may explain the broadening of the distribution of the Hamming distance. Apart from these degeneracies, the Epinions network contains many small disjoint motifs (scattered at all frequencies ν) whose flipping slightly

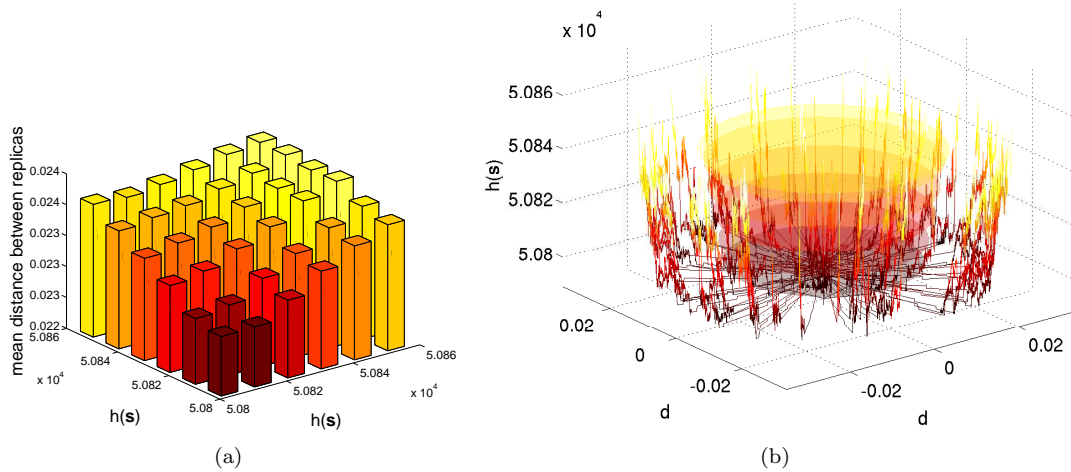


FIG. 4: Epinions. (a): Distance among replicas versus energy of a replica. The replicas of Fig. 1(a) are binned into 6 bins according to their energy, and the mean of the relative Hamming distances is computed. The vertical axis (and color code) represents the mean over bins of the relative distances. The replicas of least energy are also closer, and the distance grows regularly with the energy. (b): Sample minimal energy paths connecting the global and a local minimum. For visualization purposes, the trajectories are depicted as radially distributed according to a polar coordinate, with the global optimum placed in the origin. The vertical axis (and color code) represents the energy. The radius of the disks represent the average distance among the replicas in correspondence of the 6 bins of (a). The horizontal parts on the paths correspond to isoenergetic flips.

increases the energy. Their cumulative effect is responsible for the energy difference between replicas, which smoothly grows moving away from the ground state.

TABLE 4: Slashdot communities: nodes and edges of the 3 communities C_i identified from Fig. 3. The row *rest* denotes the complement to the C_i in \mathcal{J} . Internal edges are those connecting two nodes of the same C_i , while external nodes are those connecting a node in C_i and one in C_j (or in *rest*). The C_i are visualized in Fig. 6.

<i>community</i>	nodes		edges		
	n_{C_i}	+1 int.	-1 int.	+1 ext.	-1 ext.
C_1	855	4169	1362	6268	6587
C_2	62	11	107	143	148
C_3	58	7	197	443	1009
<i>rest</i>	51073	345510	102856	6752	7259

B. Slashdot and competing valleys of near-optimal balance.

For Slashdot, the low-energy landscape is markedly different: distinct peaks in the Hamming distances are clearly visible (Fig. 2) and are

related to the three valleys observed in the energy histogram of Fig. 1(b). In fact, a scatter plot (Fig. 5(a)) reveals that while the two valleys at higher energy are nearby also in configuration space (yellow points in Fig. 5(a)) both of them are far away from the lowest valley (cyan and magenta points). To confirm that indeed this multivalley profile is not due to undersampling of the low-energy landscape, the search for replicas in $[\delta_{\text{up}}, \delta_{\text{up}} + \epsilon]$ was performed a large number of times (compare the two values of r in Table 2). Even increasing tenfold r no new valley emerged for this network.

Looking at the spin frequencies of Slashdot (Fig. 3(b)) we can also observe a feature absent in Epinions, namely the presence of three sharp spike-like peaks for $z(\nu)$. If broad peaks usually contain small connected components, sharp, spike-like peaks in the plot of $z(\nu)$ are more likely to be associated to large connected components whose spins are simultaneously flipped in some of the replicas. A thorough analysis reveals that these three peaks correspond in fact to three connected communities characterized by an high level of internal balance, higher than with the rest of the network. In Table 4 we report some features of the corresponding subnetworks and in Fig. 6 we draw their adjacency matrices and graphs. These communities are responsible for the formation of competing valleys of near-optimal balance described

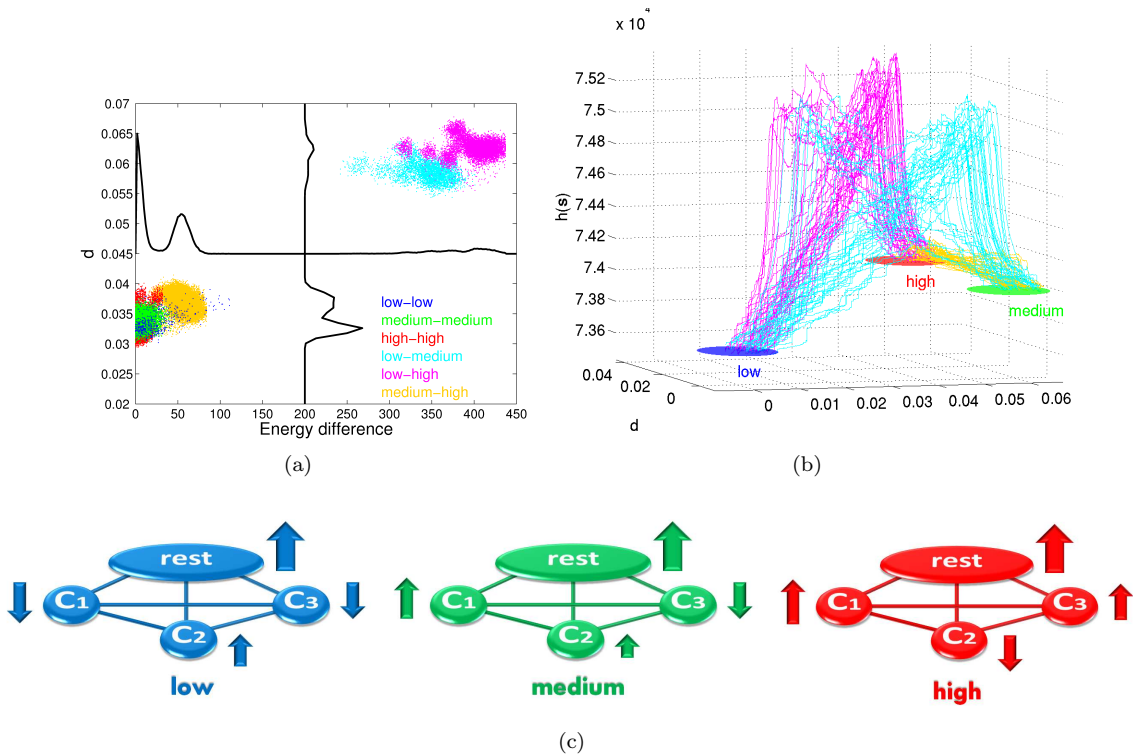


FIG. 5: Slashdot. (a): The three valleys of minima corresponding to the three peaks in Fig. 1(b) are here called *low*, *medium* and *high*, according to their energy. For each pair of minima $\mathbf{s}^{(i)}$ and $\mathbf{s}^{(j)}$, the scatter plot shows differences in energy $|h(\mathbf{s}^{(i)}) - h(\mathbf{s}^{(j)})|$ versus the relative Hamming distance d for both intra and intervalley peaks. The two valleys *medium* and *high* are near and both more distant from the *low* valley. (b): A few sample trajectories connecting minima in different valleys. The radius of the disks corresponds to the average intravalley distance among the replicas. The color code is the same as in (a). The height of the trajectories is indicative of the energy barrier between the valleys. To improve readability, the degenerate spin flips are not shown. (c): The three internally balanced communities C_i (described in Tables 4-5 and Fig. 6) and their average magnetization in the three valleys of near-optimal balance (color-code as in (a)).

earlier. In fact, while the internal arrangement of their spins is usually frozen on each valley, their relative orientation with respect to the rest of the network may change passing from one valley to another, meaning that all spins of a community are simultaneously flipped, see Fig. 5(c). Computing a few sample trajectories from one valley to another (Fig. 5(b)), we can observe the presence of an energy barrier: the paths break the internal balance of some of the communities before they are able to rearrange them again in another manner. In terms of our social networks, the formation of different valleys of near-optimal balance means the presence of possible alternative “alliances” between the majority of the users and a few internally balanced communities of users (not necessarily mutually friends among them). Fixing the relationship with one such community constraints the sign of the relationship with the other communities. Different arrangements lead to slightly different global levels of balance for the whole network.

TABLE 5: Slashdot valleys: $\langle \mathbf{s}_{C_i} \rangle_{low, medium, high}$ is the average magnetization of the community C_i (or of *rest*) in the three low-energy valleys of Fig. 5.

$\langle \mathbf{s}_{C_i} \rangle$	<i>valley</i>		
<i>community</i>	<i>low</i>	<i>medium</i>	<i>high</i>
C_1	-0.712	0.712	0.712
C_2	0.513	0.066	-0.663
C_3	-0.517	-0.517	0.506
<i>rest</i>	0.797	0.804	0.807

As already mentioned, a necessary condition for the spins of a subcommunity to have constant sign relative to each other in all low-energy replicas (which means also to be flipped simultaneously in all these near-optimal configurations) is that the community has to have a high level of balance internally, and a certain amount of frustration with respect to the rest of the network. At low energies, in fact, this favors a constant choice of spin orien-

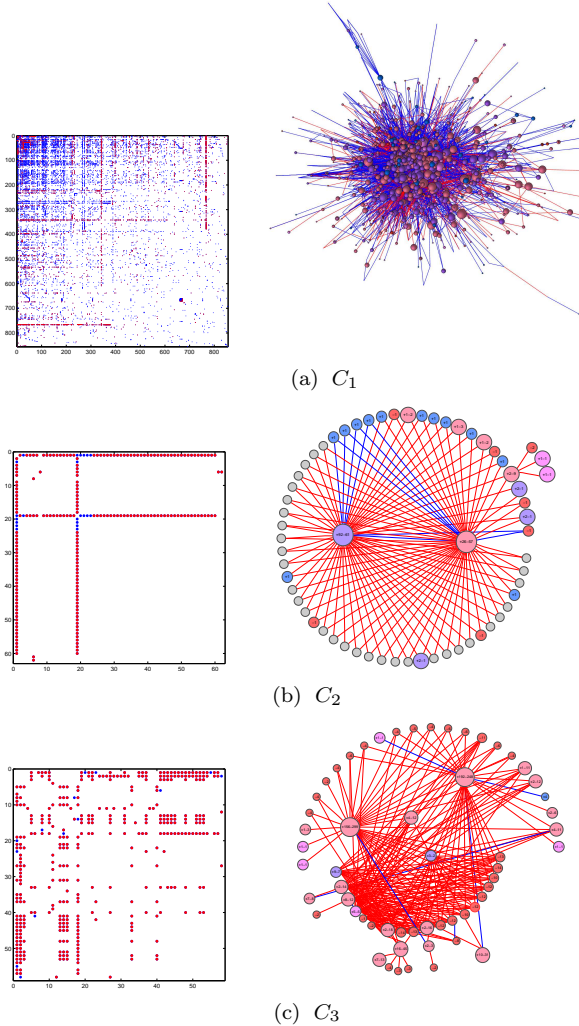


FIG. 6: Slashdot: the three internally balanced communities C_1 , C_2 and C_3 of Table 4. The adjacency matrices and the corresponding signed graphs are shown. Blue dots correspond to +1 edges, red to -1. The color-code for the nodes reflects the sum of the internal edges (blue means positive sum, red negative). The numbers inside the nodes represent the external edges (not shown for C_1).

tation within the community which can however vary from valley to valley.

Denote C_1 , C_2 and C_3 the three communities of Table 4 (and *rest* their complement in the original signed graph). By lumping together all nodes of each community, the corresponding matrix of adjacencies (blocks ordered as C_1 , C_2 , C_3 and *rest*) is the following:

$$\mathcal{A} = \begin{bmatrix} 2807 & 16 & -206 & -129 \\ 16 & -96 & -1 & -19 \\ -206 & -1 & -190 & -359 \\ -129 & -19 & -359 & 242654 \end{bmatrix} \quad (4)$$

where the original amounts of positive and negative edges are:

$$\mathcal{A}(+/-) = \begin{bmatrix} 4169/1362 & 30/14 & 21/227 & 6217/6346 \\ 30/14 & 11/107 & 0/1 & 113/132 \\ 21/227 & 0/1 & 7/197 & 422/781 \\ 6217/6346 & 113/132 & 422/781 & 345510/102856 \end{bmatrix} \quad (5)$$

Both C_2 and C_3 have a vast majority of internal negative edges, see Fig. 6 and Table 4. In correspondence of the optimal balance, the gauge transformed adjacency matrix for the communities is

$$\mathcal{A}_\sigma(+/-) = \begin{bmatrix} 5441/90 & 32/12 & 242/6 & 6726/5837 \\ 32/12 & 118/0 & 1/0 & 170/75 \\ 242/6 & 1/0 & 204/0 & 809/394 \\ 6726/5837 & 170/75 & 809/394 & 381167/67199 \end{bmatrix}$$

where it can be observed that the various C_i have very few residual negative edges (in particular C_2 and C_3 have none), meaning that indeed these communities are internally balanced. If we look at the structure of the adjacency matrices of the C_i in Fig. 6, it is easy to understand why so many negative edges disappear in the gauge transformed \mathcal{A}_σ : the negative edges are all in correspondence of the same users. This is particularly visible in the adjacency matrix of C_1 : each row/column is highly skewed towards positive or negative edges. Such skewed sign distributions are the trademark for “apparent disorder”, i.e., negative edges which can be eliminated by means of gauge transformations and hence that do not spoil global balance. In Slashdot, users with a high number of negative edges are known as trolls [19]. Trolls do not add tension to the network, as they are unanimously tagged as foes by the other users. Looking carefully at Fig. 6 it is possible to observe that the communities C_2 and C_3 , where negative edges are the vast majority, have nevertheless all positive cycles. In C_2 all cycles have length 3 and pass through the (positive) edge linking the two highly connected nodes. In C_3 , instead, cycles have length 4 and are composed of 4 negative edges.

Observing the pattern of signs (and sign flips) of the C_i in the 5557 low-energy replica of Slashdot, a high degree of regularity can be seen. Call $\mathbf{s}_{C_i}^{(j)}$ the spin configuration of the C_i community in the j -th low-energy replica, and n_{C_i} the number of spins of C_i . By construction, the relative Hamming distance within the community

$$d(\mathbf{s}_{C_i}^{(j)}, \mathbf{s}_{C_i}^{(k)}) = \left(1 - (\mathbf{s}_{C_i}^{(j)})^T \mathbf{s}_{C_i}^{(k)} / n_{C_i}\right) / 2$$

is always zero. The intervalley relative Hamming distance is shown in Fig. 5(a). The “average magnetization” within a community and within a valley is computed restricting the computation to the

\mathbf{s}_{C_i} spins and to the replicas falling into the valley. If r_{low} is the number of replicas in the *low* valley, then the average magnetization for the *low* valley is

$$\langle \mathbf{s}_{C_i} \rangle_{low} = \sum_{k \in C_i} \sum_{j \in low} s_k^{(j)} / (r_{low} n_{C_i}),$$


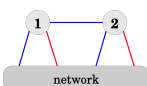
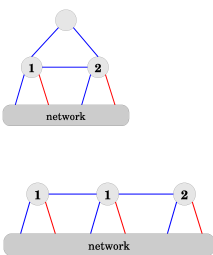
and similarly for $\langle \mathbf{s}_{C_i} \rangle_{medium}$ and $\langle \mathbf{s}_{C_i} \rangle_{high}$, $i = 1, 2, 3$. These average magnetizations are shown in Fig. 5(c) and Table 5. It can be observed that for example the spins \mathbf{s}_{C_1} are flipped passing from the *low* energy valley to the *medium/high* energy valleys. These community-wide flips do not modify the intravalley energy, but they alter the energy of the cut set between the C_i s and with the rest of the network. The increase in energy passing from one valley to another is a consequence of these changes. The cut sets between C_1 and C_2 and between C_2 and C_3 are always negligible, while the cut set between C_1 and C_3 can increase the energy considerably. This happens when $\langle \mathbf{s}_{C_1} \rangle$ and $\langle \mathbf{s}_{C_3} \rangle$ are not solidary, i.e., in the *medium* valley. This makes most of the difference in energy between *low* and *medium* valleys. In the *high* valley, instead, the magnetizations $\langle \mathbf{s}_{C_1} \rangle_{high}$ and $\langle \mathbf{s}_{C_3} \rangle_{high}$ are again aligned. However, both are flipped with respect to the *low* valley, and this reflects in the changes of energies in the cut sets relative to the *rest* of the network. Remember that *rest* contains a very large fraction of the network, and that it is highly biased towards positive magnetization (see Table 5), which is essentially uninfluenced by flipping of small communities. Nevertheless, the internal rearrangements inside *rest* due to the flips of \mathbf{s}_{C_1} and \mathbf{s}_{C_3} induce a consistent increase of frustration within *rest*. In summary, it appears that a feature instrumental to the creation of separated energy valleys is the presence of perfectly balanced communities (like C_2 , C_3 and, to a large extent, also C_1). These can be composed of friends as well as of “declared” enemies: for what concerns structural balance, the trolls of C_2 and C_3 play exactly the same role as the users having many positive edges.

V. CONCLUSION

The possibility of exploring the low-energy landscape of the equivalent Ising model of large signed social networks is an important tool for investigating some peculiarities of the social interactions which take place inside these communities. In the paper we have presented examples of what can be achieved using a large number of low energy minima for the characterization of two large-scale online networks, Epinions and Slashdot.

In spite of a similar fraction of negative edges in

TABLE 6: Isoenergetic motifs identified under the peak at $\nu = 1/2$. The values c_i^+ and c_i^- refers to the number of positive and negative edges from the i -th node to the rest of the network in the gauge transformed \mathcal{J}_σ . The motifs are classified according to their size and then to the size of the cut set with the rest of the network. In the case marked with the asterisk, one of the triangles on the top presents a frustration.

motif	c_1^+	c_1^-	c_2^+	c_2^-	Epinions	Slashdot
	1	1	-	-	1928	1469
	2	2	-	-	180	155
	3	3	-	-	40	41
	4	4	-	-	15	14
	5	5	-	-	6	6
	6	6	-	-	2	6
	7	7	-	-	1	2
	8	8	-	-		3
	9	9	-	-		2
	17	17	-	-		
24	24	-	-			
1 node	total				2172	1698
	1	1	1	1	13	5
	1	1	2	2	5	2
	1	1	3	3	2	2
	1	1	4	4	1	1
	1	1	5	5	2	
	2	2	2	2	1	1
	2	2	3	3		1
2 nodes	total				24	12
	1	1	0	0	2	4*
	2	2	0	0		2
	4	4	0	0	1	
	1	1	1	1	1	2
	1	1	2	2		1
3 nodes	total				4	9

the two networks (16.5% for Epinions and 23.9% for Slashdot) and of a similar low level of frustration (partially ferromagnetic behavior), two quite different low-energy landscapes have been identified: a single valley for Epinions and a set of three competing valleys for Slashdot. We associate this different qualitative behavior with the presence of exactly balanced (and highly connected)

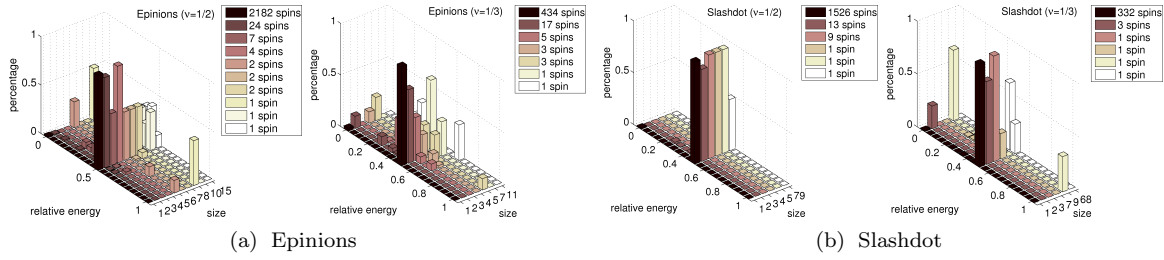

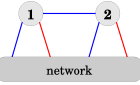
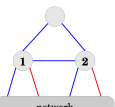


FIG. 7: Relative energy of the motifs flipping with frequencies $\nu = 1/2$ and $\nu = 1/3$. These motifs are grouped according to their size. The gray-scale (color online) represents their number. In all histograms the relative energy (energy through the cut set divided by the corresponding number of edges) is concentrated around 0.5, meaning that half of the edges are frustrated in the ground state. Therefore, the peaks at $\nu = 1/2$ and $\nu = 1/3$ contain mostly isoenergetic motifs. These are described in Table 6-7.

TABLE 7: Isoenergetic motifs identified under the peak at $\nu = 1/3$. The c_i^+ and c_i^- are the positive and negative edges of the i -th node through the cut separating a motif from the rest of the network in the ground state. In the case marked with a double asterisk the two nodes are interconnected by a negative edge.

motif	c_1^+	c_1^-	c_2^+	c_2^-	Epinions	Slashdot
 network	1	1	-	-	367	245
	2	2	-	-	51	53
	3	3	-	-	17	13
	4	4	-	-	2	6
	5	5	-	-	1	
	6	6	-	-	1	1
	9	9	-	-	1	
	15	15	-	-	1	
	17	17	-	-	1	
	20	20	-	-	1	
1 node	total				442	318
 network	1	1	1	1	2	
	1	1	2	2	1	
	1	1	3	3	2	
	2	2	3	3		1
	1	1	34	34		
	1	0	0	1	3	1
	2	1	0	1	1	
1	0	0	1		1**	
2 nodes	total				9	3
 network	1	0	0	1	1	1
	3 nodes	total				1

subcommunities, whose internal balance must be destroyed and reformed in order to pass from one valley to another. It is worth observing that these subcommunities are not necessarily composed by friends, making the determination of energy barriers like these a difficult task if one simply looks at the sign distributions on the edges.

These results represent a clear demonstration that, for complex systems like large social networks, global properties cannot be inferred by local or mean features: more sophisticated analytical and computational tools must be developed and applied.

Acknowledgments. C.A. acknowledges financial support from MIUR (Ministero dell’Istruzione, dell’Università e della Ricerca) and INdAM (Istituto Nazionale di Alta Matematica). The EU-IndiaGRID2 project (European FP7 e-Infrastructure Grant 246698) is acknowledged for the use of its grid infrastructure.

Appendix

A. Isoenergetic motifs.

Both the peaks at $\nu = 1/2$ and $\nu = 1/3$ described in Fig. 3 contain a large number of small disconnected components. For the two networks, the vast majority of these connected components are isoenergetic. In Fig. 7 these correspond to motifs having relative energy e^{rel} through the cut set equal to 0.5, where the relative energy is computed as the energy through the cut set that isolates the motif from the rest of the network, divided by the number of edges in the cut set itself:

$$e_i^{rel} = \sum_{\substack{i \in \text{motif} \\ j \notin \text{motif}}} (1 - J_{ij}s_i s_j) / (2 \sum_{i \in \text{motif}} c_i),$$

with c_i the connectivity through the cut set of the i -th node of the motif (c_i^+ and c_i^- are the corresponding numbers of $+1$ and -1 edges of c_i). Hence $e_i^{rel} = 0$ means that with respect to the ground state the i -th motif has all “satisfied” edges across the cut set, $e_i^{rel} = 0.5$ means it has 50% of frustrated edges and $e_i^{rel} = 1$ means 100% of edges frustrated. A catalog of the isoenergetic motifs under the two peaks is given in Tables 6 and 7.

While the presence of isoenergetic motifs under the $\nu = 1/2$ peak is straightforward to explain, the abundance of such motifs under the $\nu = 1/3$ peak is less obvious and requires an extra bit of investigation. On what follows we restrict ourselves to the size-1 isoenergetic motifs of Epinions. For each low energy replica σ (with the relative gauge-transformed adjacency matrix \mathcal{J}_σ) and for each node i (size-1 isoenergetic motif under either the peak at $\nu = 1/2$ or that at $\nu = 1/3$) we calculate the percentage of edges which change sign with respect to the initial matrix \mathcal{J} . This value is given by the ratio:

$$\bar{\rho}(\sigma, i) = \frac{\|\mathcal{J}_\sigma^{(i,\cdot)} - \mathcal{J}^{(i,\cdot)}\|_1}{2\|\mathcal{J}^{(i,\cdot)}\|_1},$$

where $\mathcal{J}^{(i,\cdot)}$ represents the i -th row of the matrix and $\|\cdot\|_1$ the 1-norm. We can easily attribute a plausible meaning to some of the values that this ratio can assume:

- $\bar{\rho}(\sigma, i) = 0$: the gauge transformation σ flips neither node i nor its neighbors;
- $\bar{\rho}(\sigma, i) = 1/2$: the gauge transformation σ flips only half (mostly one) of the neighbors of node i ;
- $\bar{\rho}(\sigma, i) = 1$: the gauge transformation σ flips the node i (or, less likely, all its neighbors).

In order to compare two differently populated peaks at $\nu = 1/2$ and $\nu = 1/3$, for each σ we must normalize the counts for the three cases:

$$\begin{aligned} \rho_0(\sigma) &= \frac{\#[\bar{\rho}(\sigma, i) = 0]}{\#i}; \\ \rho_{1/2}(\sigma) &= \frac{\#[\bar{\rho}(\sigma, i) = 1/2]}{\#i}; \\ \rho_1(\sigma) &= \frac{\#[\bar{\rho}(\sigma, i) = 1]}{\#i}. \end{aligned}$$

As most pairs (σ, i) are such that $\bar{\rho}(\sigma, i)$ assume the values 0, 1/2, or 1, we have the “empirical” constraint:

$$\rho_0(\sigma) + \rho_{1/2}(\sigma) + \rho_1(\sigma) \approx 1 \quad \forall \sigma. \quad (6)$$

The comparison of the values obtained for Epinions is shown in Fig. 8. For ρ_1 (bottom panel) the plots at $\nu = 1/2$ and at $\nu = 1/3$ almost totally overlap. The top plot for ρ_0 which, as already mentioned, refers to both node and neighboring nodes unchanged in a replica, instead shows a systematic difference (less frequent in $\nu = 1/3$ than in $\nu = 1/2$). From Eq. (6), such a difference is compensated by an equal but opposite difference in the middle panel for $\rho_{1/2}$. Recalling that $\rho_{1/2}$ refers to the case in which 50% of the neighboring nodes of node i are flipped, then we can conclude that the peak at $\nu = 1/3$ appears to be due, at least to some extent, to a bulk of the network which is less “rigid” than under the $\nu = 1/2$ peak and hence allows for more frequent internal rearrangements. It is worth noting that this behavior is uniform across all low-energy replicas.

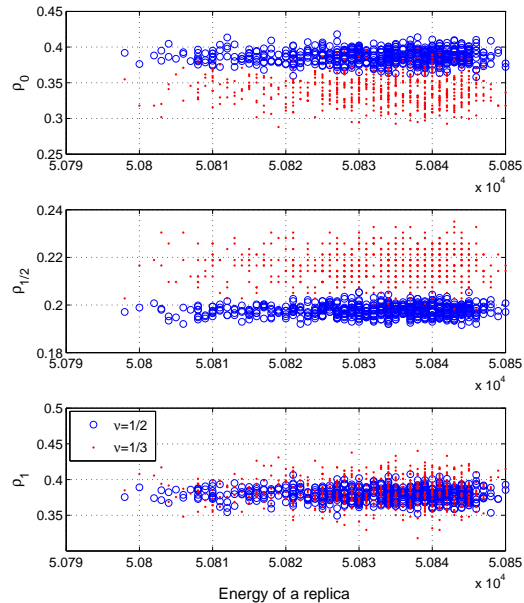


FIG. 8: Epinions: origin of the size-1 motifs for $\nu = 1/2$ and $\nu = 1/3$. Plots of the ρ_0 , $\rho_{1/2}$ and ρ_1 ratios for the 2172 isoenergetic size-1 motifs identified under the peak at $\nu = 1/2$ (circles) and for the 442 size-1 isoenergetic motifs belonging to the peak at $\nu = 1/3$ (points). The case $\rho_{1/2}$ (referring to gauge transformations that change 50% of the edges through the cut set) is more frequent under the $\nu = 1/3$ peak than under the $\nu = 1/2$ one.

-
- [1] T. Antal, P. L. Krapivsky, and S. Redner, *Phys. Rev. E* **72**, 036121 (2005).
- [2] G. Facchetti, G. Iacono, and C. Altafini, *PNAS* **108**, 20953 (2011).
- [3] S. Galam, *Physica A: Statistical and Theoretical Physics* **230**, 174 (1996).
- [4] K. Kulakowski, *Computing in Science & Engineering* **9**, 80 (2007).
- [5] S. A. Marvel, S. H. Strogatz, and J. M. Kleinberg, *Phys. Rev. Lett.* **103**, 198701 (2009).
- [6] A. Srinivasan, *Proceedings of the National Academy of Sciences* **108**, 1751 (2011).
- [7] M. Szell, R. Lambiotte, and S. Thurner, *Proceedings of the National Academy of Sciences* **107**, 13636 (2010).
- [8] F. Heider, *Journal of Psychology* **21**, 107 (1946).
- [9] D. Cartwright and F. Harary, *Psychological Review* **63**, 277 (1956).
- [10] G. Toulouse, *Communications on Physics* **2**, 115 (1977).
- [11] D. C. Mattis, *Physics Letters A* **56**, 421 (1976).
- [12] O. C. Martin, in *New Optimization Algorithms in Physics*, edited by A. K. Hartmann and H. Rieger (Wiley-VCH, 2005).
- [13] A. K. Hartmann and H. Rieger, *Optimization Algorithms in Physics* (Wiley-VCH, 2001).
- [14] J. Leskovec, D. Huttenlocher, and J. Kleinberg, in *ACM SIGCHI Conference on Human Factors in Computing Systems* (ACM SIGCHI, Atlanta, GA, 2010).
- [15] G. Iacono, F. Ramezani, N. Soranzo, and C. Altafini, *IET Systems Biology* **4**, 223 (2010).
- [16] M. Mezard, G. Parisi, and M. Virasoro, *Spin glass theory and beyond* (World Scientific, 1986).
- [17] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, in *Proceedings of the 13th international conference on World Wide Web* (2004), WWW '04, pp. 403–412.
- [18] C. A. Lampe, E. Johnston, and P. Resnick, in *Proceedings of the SIGCHI conference on Human factors in computing systems* (2007), CHI '07, pp. 1253–1262.
- [19] J. Kunegis, A. Lommatzsch, and C. Bauckhage, in *18th International World Wide Web Conference* (18th Int WWW Conf., 2009), pp. 741–741.
- [20] K. Binder and A. P. Young, *Rev. Mod. Phys.* **58**, 801 (1986).
- [21] S. Kirkpatrick, *Phys. Rev. B* **16**, 4630 (1977).