Balanced and fragmented phases in societies with homophily and social balance

Tuan Minh Pham,1,2 Andrew C. Alexander,3 Jan Korbel,1,2 Rudolf Hanel,1,2 and Stefan Thurner1,2,4

1Section for the Science of Complex Systems, CeMSIIS, Medical University of Vienna, Spitalgasse 23, A-1090, Vienna, Austria
2Complexity Science Hub, Vienna Josefstädterstrasse 39, A-1090 Vienna, Austria
3Department of Mathematics, Princeton University, NJ 08544, Princeton, USA
4Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

(Dated: December 22, 2020)

Recent attempts to understand the origin of social fragmentation are based on spin models which include terms accounting for two social phenomena: homophily—the tendency for people with similar opinions to establish positive relations—and social balance—the tendency for people to establish balanced triadic relations. Spins represent attribute vectors that encode multiple (binary) opinions of individuals and social interactions between individuals can be positive or negative. Recent work suggests that large systems of $N \gg 1$ individuals never reach a balanced state (where unbalanced triads with one or three hostile links remain), provided the number of attributes for each agent is less than $O(N^2)$ [Phys. Rev. Lett. 125, 078302]. Here we show that this statement is overly restrictive. Within a Hamiltonian framework that minimizes individuals’ social stress, we demonstrate that stationary, balanced, but fragmented states can be reached for any number of attributes, if, in addition to homophily, individuals take into account a significant fraction, $q$, of their triadic relations. Above a critical value $q_c$, balanced states result. This result also holds for sparse realistic social networks. Finally, in the limit of small $q$, our result agrees with that of [Phys. Rev. Lett. 125, 078302].

The concept of so-called filter bubbles captures the fragmentation of society into isolated groups of people who trust each other, but clearly distinguish themselves from “other”. Opinions tend to align within groups and diverge between them. Interest in this process of social disintegration, started by Durkheim [1], has experienced a recent boost, fuelled by the availability of modern communication technologies. The extent to which societies fragment depends largely on the interplay of two basic mechanisms that drive social interactions: homophily and structural balance. Homophily is based on the “principle” that individuals with similar opinions tend to become friends (“similarity breeds connection” [2]); indeed, like-minded individuals often form homogeneous structures in society [3]. Structural balance, first described by Heider [4], is the tendency of balanced triads to be over-represented in societies. A triad of individuals is balanced if all three individuals are mutual friends (friend of my friend is my friend) or if two friends have a mutual enemy (enemy of my enemy is my friend). Structural balance has been investigated by social scientists for a long time [5–7] and, more recently, by physicists and network scientists [8–18]. Recent contributions study the dynamics on balanced networks [19–21], the co-evolution of opinions and signed networks [22–35], and generalizations of the concept of structural balance [36]. For an overview, see [37, 38]. A general survey of statistical physics methods applied to opinion dynamics is found in [39, 40].

Among the various approaches towards social fragmentation, the minimization of social stress captured by a so-called social stress Hamiltonian has become particularly relevant. Here the opinion of individual $i$ is denoted by $s_i$ and the relation between $i$ and $j$ by $J_{ij}$ (positive or negative). Stress arises from a homophily-related term, $-\sum_{(i,j)} J_{ij} s_i s_j$, or a term reflecting social balance, $-\sum_{(i,j,k)} J_{ij} J_{jk} J_{ki}$. The former is similar in structure to the Edwards-Anderson spin-glass model [41], the latter to the Baxter-Wu three-spin interaction model [42]. These two terms have been studied separately, homophily in [43–46] and Heider balance in [47–49]; only recently were both terms integrated into a single Hamiltonian [35].

The issue of social cohesion and organization was recently studied in an attribute-based local triad dynamics model (ABLTD) [50]. There, each agent has binary opinions on $G$ attributes. If two agents agree on more attributes than they disagree on, they become friends (positive link). Agents tend to change their attributes to reduce stress in unbalanced triads. The paper showed that given a system of $N$ agents, as $N \to \infty$, the so-called “paradise state”, where all agents are friends of each other, is never reached unless the number of attributes $G$ scales as $O(N^\gamma)$, for $\gamma \geq 2$. Instead, the society remains in a stationary unbalanced state with an equal number of balanced and unbalanced triads. This means that for realistic situations, where $N$ is large and the number of attributes $G$ remains relatively small, it is impossible to reach social balance, let alone the paradise state. This statement is to some extent, contrary to empirical findings that societies are balanced to a high degree; see e.g. recent work on large scale studies [16, 17].

In this letter, motivated by the ABLTD, we propose an individual-stress-based model that takes into account the homophily effect between adjacent individuals and structural balance within a local neighborhood. The latter consists of the subset of the most relevant triads to an individual, i.e. those that involve its closest friends.
and greatest enemies. The ratio of relevant triads to the
total number of triangles the individuals belong to
determines whether society fragments or remains cohesive.
With the help of simulations on a regular network, we
show that there exists a critical size of the local neighbor-
hood above which society fragments, yet stays balanced.
We discuss the relation of the presented model to both,
the ABLDT model [50] and the social stress Hamiltonian
approach [35]. For appropriate parameter choices, both
models can be shown to correspond to special cases of
the presented model.

**Local social stress model.** Consider a society of \( N \)
individuals. Each individual \( i \) has binary opinions on \( G \)
issues, characterized by an attribute vector, \( \mathbf{A}_i = \{a^\ell_i\} \),
where \( a^\ell_i \in \{-1, +1\}; \ell \in 1, \ldots, G \). Further, \( i \) has
relations to \( k_i \) other individuals in a social network.
Network topology does not change over time. Following
[50], the relation between two agents \( i \) and \( j \) is deter-
mined by the sign of their distance in attribute space:
\( J_{ij} = \text{sign}(\mathbf{A}_i \cdot \mathbf{A}_j) \), where the dot denotes the scalar
product. \( J_{ij} = 1 \) indicates friendship, \( J_{ij} = -1 \) enmity.
Each agent \( i \) has a social stress level, \( H^{(i)}(\mathbf{A}) \), defined as
\[
H^{(i)}(\mathbf{A}) = -1 \frac{1}{G} \sum_j J_{ij} \mathbf{A}_i \cdot \mathbf{A}_j - \sum_{(j,k) \in Q_i} J_{ij} J_{jk} J_{ki} \cdot (1)
\]
The first sum extends over all \( k_i \) neighbours of \( i \), while
the second is restricted to \( Q_i \) (out of \( N_i^A \) possible) triads
that node \( i \) belongs to. The notation \( (j,k) \in Q_i \) means to
sum over all pairs of \( j \) and \( k \) which, together with \( i \), form
the \( Q_i \) triads. These are chosen at each step of the dy-
namics (see below). \( Q_i \) represents the number of triads \( i \)
would like to have socially balanced – \( i \)’s relevant neigh-
borhood. The factor \( 1/G \) ensures that contributions from
any link towards \( H^{(i)}(\mathbf{A}) \) do not diverge in the limit \( G \to \infty \).
Assuming agents try to minimize their individual social
stress over time, we implement the following dynamics:\[1, 2]:

1. **Initialize.** Each node is assigned an opinion vector,
   \( \mathbf{A}_i \), whose components are randomly chosen to be
   1 or −1 with equal probability. Every node has the
   same degree, \( k_i = K \), and is connected to its neigh-
bours in a regular way, forming a ring topology.
   The topology is fixed over time. For any pair of con-
   nected agents, \( i \) and \( j \), we set \( J_{ij} = \text{sign}(\mathbf{A}_i \cdot \mathbf{A}_j) \).

2. **Update.** (i) Pick a node \( i \) randomly and choose \( Q_i \)
of its triads, also randomly. Compute \( H^{(i)}(\mathbf{A}) \). In the
current state its value is \( \mathcal{H} \).
   (ii) Flip one of \( i \)’s attributes at random. Let \( \tilde{\mathbf{A}}_i \)
   be its new opinion vector. For each of the chosen
   triads, the weights of the two links adjacent to \( i \) are
   recomputed as \( \tilde{J}_{ij} = \text{sign}(\tilde{\mathbf{A}}_i \cdot \mathbf{A}_j) \). \( J \) is the new
   matrix.
   (iii) Compute the new stress \( \tilde{\mathcal{H}} \) using \( \tilde{J} \). The
   change in stress is \( \Delta H^{(i)} = \mathcal{H} - \tilde{\mathcal{H}} \).
   (iv) Update the system \( \mathbf{A}_i \rightarrow \tilde{\mathbf{A}}_i \), and \( J_{ij} \rightarrow \tilde{J}_{ij} \)
   with probability, \( \min \left\{ e^{-\Delta H^{(i)}} \cdot 1 \right\} \), otherwise leave
   it unchanged. This stochastic rule means that
   agents are not always rational and might choose
to increase their stress.

3. Continue with the next timestep.

Figure 1 illustrates an update where by changing one
opinion, agent \( i \) becomes an enemy of \( j \), but the chosen
triad, \((ijk)\), becomes balanced. If \( Q = 1 \), this decreases
\( i \)’s social stress from \( H^{(i)} = -5/3 \) to \( H^{(i)} = -3 \). If more
triads are chosen, \( Q > 1 \), this flip leads to a stress
increase and is less likely accepted.

According to the dynamical rule, the maximum num-
ber of links which may change their signs due to an opinion
update depends on \( Q \). When \( Q = 1 \), at most two
links can flip, while for \( Q \geq K \) all the \( K \) links might.
The change in stress for agent \( i \), \( \Delta H^{(i)}(\mathbf{A}) \), given that
attribute \( a^\ell_i \) flips, can be written as
\[
\Delta H^{(i)} = 2 \frac{G}{J} \sum_{(j|J_{ij} = J_{ij})} J_{ij} a^{\ell_i}_i a^{\ell_j}_j + \sum_{(j,k) \in Q} \Delta_{jk}, \quad (2)
\]
where \( j|\tilde{J}_{ij} = J_{ij} \) means to sum over those \( j \) (neigh-
bours of \( i \)) whose relations to \( i \) remain unchanged and

---

\[1\] By definition, \( N_i^A \equiv c_i k_i (k_i - 1) / 2 \), where \( c_i \) is the local clus-
tering coefficient. \( k_i \) is the degree.

\[2\] The relevance of the first term in the model dynamics is dis-
cussed in the Supplemental Material.
where $n_+$ and $n_-$ are the number of balanced and unbalanced triangles, respectively. $f = 1$ means that all triangles are balanced, $f < 1$ signals the presence of unbalanced triangles. Even though $f$ could be negative, this situation is never observed in simulations. A fundamental result in structural balance theory states that if the network can be partitioned into clusters, within which all links are positive and between which links are exclusively negative, then all triangles are balanced, and $f = 1$ [52]. Note that $f = 1$ is a necessary, but not sufficient condition for such a partition on sparse networks. We therefore call a society balanced and fragmented if all of its constituent triads are balanced and clusters of positive links are separated from each other by negative links.

**Results.** We first run the dynamics on a triangular lattice for $N = 100$ with free boundary conditions to check whether clusters of positive links can be formed. To this end, we set $Q = 2$ – i.e. only two triads per individual are chosen in each update. In Fig. 2 we show a snapshot of the signed social network after reaching a stationary state. For a small value of $G = 3$, the formation of several clusters of positive (red) links is clearly seen. Negative links are shown in light blue. This configuration belongs to the fragmented phase that shows a high degree of balance, $f \sim 0.9$. For more attributes, $G = 99$, a single cluster emerges that percolates the lattice. This is a realization in the cohesive phase that is clearly not balanced, and indeed $f \sim 0$.

Next, we run the simulation on a regular ring network for $N = 400$, where every node has a degree of $K = 32$ neighbors. Figure 3 (a) shows a phase diagram of the order parameter, $f$, which indicates a transition from an unbalanced (cohesive) to a balanced (fragmented) society. For any given $G$, this transition occurs as $q = Q/N^\Delta$ passes a threshold $q_c$. The existence of a critical $q_c$ demonstrates the importance of Heider’s balance term in driving a society towards fragmentation: if many triangles are taken into account, society becomes fragmented. For a wide range of $G$, $q > 1/5$ clearly suffices to be in the balanced and fragmented phase. Also, $q_c$ increases with growing $G$, indicating that when more issues become relevant for homophily, the chance for fragmentation lowers. Note that the situation resembles non-equilibrium in the sense that in the cohesive phase the system only reaches quasi-stationary unbalanced states which, due to fluctuations in finite-sized systems, eventually become highly balanced after a very long time. Since the presented model is stochastic, these final states are not necessarily frozen (absorbing). This means that a small number of unbalanced triads still fluctuate over time. The transition is presumably first-order, as a region of bi-stability is numerically observed where the order parameter can be $f \sim 0$ or $f \sim 1$. We checked the hypothesis that for a fixed $q$, the fragmented phase can be reached if the network degree exceeds a critical value, $K_c$; see Fig 3 (b).
The transition becomes sharper at higher $K$.

Limit of small $Q$. In the limit $Q = 1$ and $G \to \infty$, the society is expected to reach an unbalanced stationary state. This state has similar network observables (order parameter, $f$, and the fraction of positive links, $\rho_+$) as those obtained in the ABLTD model [50], although it differs in both, the notion of structural balance and the detailed update dynamics. While we consider two types of balanced triangles (Heider’s definition), [50] considers only those with three positive links. Also, [50] introduces an exogenous bias for flipping negative ties, $p$. This leads to a “paradise state” (all links positive) for $p \geq 1/2$. There is no such bias in our model and positive links can switch to negative, and vice versa, with the same probability. In the stationary state, the number of positive links is therefore approximately equal to that of negative ones. We can show this in a mean-field approach for fully-connected networks; see Supplemental Material. Our approach is motivated by the observation that the two links of a chosen triad are not likely to be flipped simultaneously, as $G \to \infty$. Instead, only one of them would be flipped at every update. A set of rate equations for triads of different types then is derived. Its steady state solution is shown to be $f^{(st)} = 0$ and $\rho^{(st)}_+ = 1/2$, in full agreement with [50] for $p = 1/3$, and for an arbitrary number of attributes.

Limit of large $Q$. Another interesting limit is when $Q \to N^\Delta$. In this case, one can compare the model with the Hamiltonian approach used in [35], in which the contribution of all $N \times N^\Delta / 3$ triangles, weighted by a coupling $g$, is taken into account:

$$H \equiv -\frac{1}{2G} \sum_{(i,j)} J_{ij} A_i \cdot A_j - g \sum_{(i,j,k)} J_{ij} J_{jk} J_{ki}. \quad (4)$$

Here the first sum extends over all connected pairs, the second over all triangles. In Eq. (4), in contrast to the model presented here, $J_{ij}$ are random dynamical variables that co-evolve with, but are not strictly determined by the opinion vectors. The detailed updating procedure of [35], which aims at minimizing $H$, is described in the Supplemental Material. Despite the differences in the concrete update dynamics, for a large enough $Q \geq Q_{MF}$, the two models are expected to yield similar results if $g$ is related to $Q$ by $g = \alpha Q / N^\Delta$, for some constant $\alpha$. Here the main idea is that for sufficiently large $Q$, individuals’ actions have a similar outcome regardless of their knowledge of the total stress $\bar{H}$ in the society. Figure 4 shows the comparison for $\alpha = 1$. The curve of the presented model indeed crosses that of the model given by Eq. (4) at $q_1 \geq Q_{MF} / N^\Delta \approx 0.133$ for $G = 23$ in (a), and at $q_2 \geq 1/6$, for $G = 99$ in (b), where the coupling, $g$, is chosen to be equal to $q_1$ in (a) and $q_2$ in (b), respectively. The transition in both models also shows that the fragmented and balanced phase necessarily exists if the effect of social balance is sufficiently large with respect to that of homophily, regardless of the stress and the details of the dynamics under consideration.

We showed that society can achieve structurally balanced states if individuals’ opinions co-evolve with their social links so as to minimize their individual social stress. The two main structure-relevant principles in social dynamics – homophily and structural balance – are simultaneously taken into account by the model. The parameter $G$ controls the dimension of the opinion vectors relevant for homophily, and $Q$, specifies how many triangles individuals actually consider for their local social balance. The interplay between the two mechanisms results in a nontrivial phase diagram showing an abrupt change in patterns of social structure. We find two regimes: fragmentation and cohesion. In the former, society is fragmented into locally cooperative clusters of agents who are linked positively within and negatively between clusters. In the latter, globally percolating cooperation is realized by the existence of a large connected component of positively linked agents. The transition between the regimes is numerically observed at a critical fraction of the considered triangles, $q_c$, illustrating the main message of the paper: If people follow too many friends, society is more likely to become fragmented. This message is robust with respect to the change of the network connectivity; for a

FIG. 3. Order parameter, $f$, (a) as a function of $q$ and $G$ for $K = 32$, and (b) as a function of $K$ and $G$ with $q = 1/3$. Results are averaged over 100 runs on a regular ring network with $N = 400$.

FIG. 4. Limit of large $Q$. Comparison of the presented model in Eq. (1) and the one in Eq. (4) that resembles [35]. The coupling, $g$, is chosen to be $g = q \equiv Q / N^\Delta$, where $Q$ is the number of actually updated triads; $N^\Delta$ is the total number of triads. Results are averaged over 100 independent runs for $N = 100$, $K = 32$, $N^\Delta = 300$, and $G = 23$ (a); $G = 99$ (b).
fixed value of $q$, we see that the higher the degree, the more likely the society fragments.

In the fragmented phase, most of the triads are found to be balanced, which is seemingly in contrast with the message of [50]. However, there is no contradiction; the two models can be related. If only one triangle is considered, the presented model corresponds to the particular case of [50] where there is no bias toward friendships. For this case both indeed have the same stationary value of network observables. Note that if an equivalent to their bias term, $p$, is introduced to Eq. (1) in the form $h(\sum_{ij} 1 - J_{ij})$ with external field strength, $h$, then the “paradise state” can be reached for sufficiently large $h$.

For large $Q$, the model can be related to a recent co-evolutionary model [35] which corresponds to minimizing the Hamiltonian in Eq. (4). By choosing the coupling, $g = Q/N^\Delta$, we show that there exists a value $Q_{MF}$ beyond which the two approaches produce very similar results. A widespread view in the literature is that fragmentation could emerge from collaborative efforts of individuals to minimize overall stress as local minima of this stress [43, 48, 49]. The existence of $Q_{MF}$, however, suggests that this view is only a good approximation if individuals keep a large fraction of their local triads balanced. Previous works studying social fragmentation under the joint effects of homophily and social balance have been in only partial agreement with Heider’s theory. The model presented here shows the possibility of social fragmentation being fully consistent with social balance theory.

This work was supported in part by Austrian Science Fund FWF under P 29252 and P 29032, and by the Austrian Science Promotion Agency, FFG project under 857136. Simulations were carried out in part at the Vienna Scientific Cluster. AA acknowledges an internship at CSH.

[1] É. Durkheim, *De la division du travail social: étude sur l’organisation des sociétés supérieures* (Alcan, 1893).
[2] M. McPherson, L. Smith-Lovin, and J. M. Cook, Birds of a feather: Homophily in social networks, Annu. Rev. of Sociol. 27, 415 (2001).
[3] G. A. Huber and N. Malhotra, Political homophily in social relationships: Evidence from online dating behavior, J. of Politics 79, 269 (2017).
[4] F. Heider, Attitudes and cognitive organization, J. Psychol. 21, 107 (1946).
[5] F. Harary, On the notion of balance of a signed graph, Michigan Math. J. 2, 143 (1953).
[6] D. Cartwright and F. Harary, Structural balance: a generalization of Heider’s theory, Psychol. Rev. 63, 277 (1956).
[7] J. A. Davis, Clustering and structural balance in graphs, Hum. Relat. 20, 181 (1967).
[8] K. Kulakowski, P. Gawronski, and P. Gronek, The Heider balance: A continuous approach, Int. J. Mod. Phys. C 16, 707 (2005).
[9] T. Antal, P. L. Krapivsky, and S. Redner, Dynamics of social balance on networks, Phys. Rev. E 72, 036121 (2005).
[10] T. Antal, P. Krapivsky, and S. Redner, Social balance on networks: The dynamics of friendship and enmity, Physica D 224, 130 (2006).
[11] F. Radicchi, D. Vilone, and H. Meyer-Ortmanns, Universality class of triad dynamics on a triangular lattice, Phys. Rev. E 75, 021118 (2007).
[12] S. A. Marvel, S. H. Strogatz, and J. M. Kleinberg, Energy landscape of social balance, Phys. Rev. Lett. 103, 198701 (2009).
[13] V. A. Traag, P. Van Dooren, and P. De Leenheer, Dynamical models explaining social balance and evolution of cooperation, PLoS One 8, 1 (2013).
[14] N. P. Hummon and P. Doreian, Some dynamics of social balance processes: bringing Heider back into balance theory, Soc. Networks 25, 17 (2003).
[15] P. Abell and M. Ludwig, Structural balance: A dynamic perspective, J. of Math. Sociol. 33, 129 (2009).
[16] M. Szell, R. Lambiotte, and S. Thurner, Multirelational organization of large-scale social networks in an online world, Proc. Nat. Acad. Sci. USA 107, 13636 (2010).
[17] J. Leskovec, D. Huttenlocher, and J. Kleinberg, Signed networks in social media, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’10 (ACM, New York, NY, USA, 2010) pp. 1361–1370.
[18] M. Sadilek, P. Klimmek, and S. Thurner, Asocial balance—how your friends determine your enemies: understanding the co-evolution of friendship and enmity interactions in a virtual world, J. Comput. Soc. Sci. 1, 227 (2018).
[19] C. Altafini, Dynamics of opinion forming in structurally balanced social networks, PLoS One 7, 1 (2012).
[20] A. V. Proskurnikov, A. S. Matveev, and M. Cao, Opinion dynamics in social networks with hostile camps: Consensus vs. polarization, IEEE Trans. Automat. Contr. 61, 1524 (2016).
[21] P. Cisneros-Velarde, N. E. Friedkin, A. V. Proskurnikov, and F. Bullo, Structural balance via gradient flows over signed graphs, IEEE Trans. Automat. Contr., 1 (2020).
[22] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Modeling echo chambers and polarization dynamics in social networks, Phys. Rev. Lett. 124, 048301 (2020).
[23] J. Gross and C. K. W. De Dreu, The rise and fall of cooperation through reputation and group polarization, Nat. Com. 10, 776 (2019).
[24] Z. Gao and Y. Wang, The structural balance analysis of complex dynamical networks based on nodes’ dynamical couplings, PLoS One 13, 1 (2018).
[25] Y. Chen, L. Chen, X. Sun, K. Zhang, J. Zhang, and P. Li, Coevolutionary dynamics of opinion propagation and social balance: The key role of small-worldness, Eur. Phys. J. B 87, 62 (2014).
[26] H. Deng, P. Abell, O. Engel, J. Wu, and Y. Tan, The influence of structural balance and homophily/heterophobia on the adjustment of random complete signed networks, Soc. Networks 44, 190 (2016).
[27] R. Singh, S. Dasgupta, and S. Sinha, Extreme variability in convergence to structural balance in frustrated dynamical systems, Europhys. Lett. (EPL) 105, 10003 (2014).
[28] P. Singh, S. Sreenivasan, B. K. Szymanski, and G. Korniss, Competing effects of social balance and influence, Phys. Rev. E 93, 042306 (2016).
[29] M. W. Macy, J. A. Kitts, A. Flache, and S. Benard, Dynamic social network modeling and analysis: Workshop summary and papers (The National Academies Press, Washington, DC, 2003) Chap. Polarization in Dynamic Networks: A Hopfield Model of Emergent Structure, p. 162.
[30] A. Parravano, A. Andina-Díaz, and M. A. Meléndez-Jiménez, Bounded confidence under preferential flip: A coupled dynamics of structural balance and opinions, PLoS One 11, 1 (2016).
[31] F. Aguiar and A. Parravano, Tolerating the intolerant: Homophily, intolerance, and segregation in social balanced networks, J. Confl. Res. 59, 29 (2015).
[32] I. Agbanusi and J. Bronski, Emergence of balance from a model of social dynamics, SIAM J. Appl. Math. 78, 193 (2018).
[33] M. Saeedian, M. San Miguel, and R. Toral, Absorbing phase transition in the coupled dynamics of node and link states in random networks, Sci. Rep. 9, 9726 (2019).
[34] M. Saeedian, M. S. Miguel, and R. Toral, Absorbing-state transition in a coevolution model with node and link states in an adaptive network: network fragmentation transition at criticality, New J. Phys. 22, 113001 (2020).
[35] T. M. Pham, I. Kondor, R. Hanel, and S. Thurner, The effect of social balance on social fragmentation, J. R. Soc. Interface 17, 20200752 (2020).
[36] A. Kirkley, G. T. Cantwell, and M. E. J. Newman, Balance in signed networks, Phys. Rev. E 99, 012320 (2019).
[37] X. Zheng, D. Zeng, and F.-Y. Wang, Social balance in signed networks, Inform. Syst. Front. 17, 1077 (2015).
[38] C. M. Rawlings and N. E. Friedkin, The structural balance theory of sentiment networks: Elaboration and test, Am. J. Sociol. 123, 510 (2017).
[39] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics, Rev. Mod. Phys. 81, 591 (2009).
[40] M. Galesic and D. Stein, Statistical physics models of belief dynamics: Theory and empirical tests, Physica A 519, 275 (2019).
[41] S. F. Edwards and P. W. Anderson, Theory of spin glasses, J. Phys. F 5, 965 (1975).
[42] R. Baxter and F. Wu, Exact solution of an ising model with three-spin interactions on a triangular lattice, Phys. Rev. Lett. 31, 1294 (1973).
[43] R. Axelrod and D. S. Bennett, A landscape theory of aggregation, British Journal of Political Science 23, 211 (1993).
[44] S. Galam, Fragmentation versus stability in bimodal coalitions, Physica A 230, 174 (1996).
[45] G. Vinogradova and S. Galam, Rational instability in the natural coalition forming, Physica A 392, 6025 (2013).
[46] G. Facchetti, G. Iacono, and C. Altafini, Computing global structural balance in large-scale signed social networks, Proc. Nat. Acad. Sci. USA 108, 20953 (2011).
[47] S. A. Marvel, S. H. Strogatz, and J. M. Kleinberg, Energy landscape of social balance, Phys. Rev. Lett. 103, 198701 (2009).
[48] F. Rabbani, A. H. Shirazi, and G. R. Jafari, Mean-field solution of structural balance dynamics in nonzero temperature, Phys. Rev. E 99, 062302 (2019).
[49] A. M. Belaza, K. Hoefman, J. Ryckebusch, A. Bramson, M. van den Heuvel, and K. Schoors, Statistical physics of balance theory, PLoS One 12, 1 (2017).
[50] P. J. Górski, K. Bochenina, J. A. Holyst, and R. M. D’Souza, Homophily based on few attributes can impede structural balance, Phys. Rev. Lett. 125, 078302 (2020).
[51] D. Hebb, The Organization of Behavior: A Neuropsychological Theory, A Wiley book in clinical psychology (Wiley, 1949).
[52] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications (Cambridge University Press, 1994).
SUPPLEMENTAL MATERIAL

The steady state of the stress dynamics in the limit $Q = 1$, $G \to \infty$ for a fully-connected network

![Diagram of different types of triads: $\Delta_k$, with $k = 0, 1, 2, 3$ being the number of negative links. In the limit $G \to \infty$, it is very unlikely, for example, that two positive links of the $\Delta_1$ triad attached to $i$ will change their signs at the same time if one of $i$’s attribute flips, resulting in a $\Delta_3$ triad.](image)

SM Fig. 1. Four different types of triads: $\Delta_k$, with $k = 0, 1, 2, 3$ being the number of negative links. In the limit $G \to \infty$, it is very unlikely, for example, that two positive links of the $\Delta_1$ triad attached to $i$ will change their signs at the same time if one of $i$’s attribute flips, resulting in a $\Delta_3$ triad.

When only one triad (Q = 1) containing node $i$ is picked in an update, two links incident to $i$, say $J_{ij}$ and $J_{ik}$, are to be redefined according to $J_{ij} = \text{sign}(A_{ij} \cdot A_{j})$ and $J_{ik} = \text{sign}(A_{ik} \cdot A_{k})$. From this definition, a change in the link $J_{ij}$ from positive to negative can happen if, and only if, before the update there are $(G + 1)/2$ common attributes between $i$ and $j$ and then one among them is selected to flip. This may happen with probability $a_+$. We can estimate

$$a_+ = \frac{1}{2^{G-1}} \left( \frac{G}{(G + 1)/2} \right) \frac{G + 1}{2^G},$$

assuming $i$ can have any of the $2^{G-1}$ possible states compatible with that of $j$. Similarly, a previously negative link can become positive with $a_- = a_+$. This reflects the fact that the process of a link flip is unbiased in our implementation. Note that

$$a_\pm \propto \frac{1}{\sqrt{G}}, \quad G \to \infty.$$

Let $\rho_+$ be the fraction of positive links. If one assumes that the link $J_{ij}$ can be positive with probability $\rho_+$ or negative with probability $1 - \rho_+$, then the probabilities that it changes from $+ \to -$ and from $- \to +$, $\pi(+-\to-)$ and $\pi(-+\to+)$, respectively, are given by

$$\pi(+-\to-) = \rho_+ a_+, \quad \pi(-+\to+) = (1 - \rho_+) a_-$$

Among the three possible cases – namely, exactly 0, 1, or 2 of the links change sign – in the limit $G \to \infty$, we can neglect the case where both of them change at the same time as this scenario happens with vanishing probability. This means that in our treatment we consider only those update events during which at most one link can switch.

We formulate the set of differential equations describing the time evolution of triads in a fully-connected network, following Antal et al. (Phys. Rev. E72, 036121). Let $\Delta_k$ be as defined in SM Fig. 1. The corresponding density of $\Delta_k$ triads is defined by $n_k = N_k/M$, where $N_k$ is the total number of triads of type $\Delta_k$ and $M \equiv \sum_k N_k = \binom{N}{3}$ is the total number of triads in a fully-connected network. We first define two variables: the density of triads of type $\Delta_k$ that are attached to a positive link, $n_k^+$,

$$n_k^+ = \frac{(3 - k)n_k}{3n_0 + 2n_1 + n_2},$$

and the density of triads of type $\Delta_k$ that are attached to a negative link, $n_k^-$,

$$n_k^- = \frac{kn_k}{n_1 + 2n_2 + 3n_3}.$$

Using $n_k^+$ and $n_k^-$, the set of ODEs can be written as

$$\begin{align*}
\dot{n}_0 &= \pi(-+\to-)n_1 - \pi(+\to-)n_0^+ \\
\dot{n}_1 &= \pi(+-\to-)n_0 + \pi(-\to+)n_1^+ - \pi(-+\to-)n_1^- - \pi(+\to-)n_1^+ \\
\dot{n}_2 &= \pi(-+\to-)n_2 + \pi(-\to+)n_3 - \pi(-+\to-)n_2^- - \pi(+\to-)n_2^+ \\
\dot{n}_3 &= \pi(+-\to-)n_3^- - \pi(+\to-)n_3^+.
\end{align*}$$

The stationary state is defined by setting the left sides of Eqs. $(*)$ to zero and requiring $\pi(+\to-) = \pi(-+\to)$ for a fixed density of positive link. This implies

$$\begin{align*}
n_0^* &= n_1, \quad n_1^* = n_2, \quad n_2^* = n_3 \\
\rho_+^* a_+ &= (1 - \rho_-^*) a_- \quad (***)
\end{align*}$$

As within our approximation scheme, $a_+ = a_-$, the steady state solution to Eqs. $(***)$ is $\rho_+^* = 1/2$. A consequence of this is the density of triangles type $\Delta_k$ is given by

$$n_k = \binom{N}{3}/8, \quad G \to \infty$$

for $G = O(N^2)$, $G \to \infty$ and for $p = 1/3$, $G$ arbitrary, where $p$ is the tendency towards friendship in their paper. One would need to introduce a $p$ dependence to the expression of $a_+$ and $a_-$ to recover the upper ($p$-dependent) part of the phase diagram reported there.

Effect of the first term in Eq. (1)

Here we check what happens to the dynamics if the first term in Eq. (1) is switched off. We find in our simulation of finite size systems without the homophily term that balanced states can still be reached due to large fluctuations in the active phase. These fluctuations, however, become smaller as $N \to \infty$, resulting in an unbalanced quasi-stationary situation. The balanced state hence may not be observed within finite time. This shows that the effects of both terms are necessary for the convergence towards the balanced state (the sufficient condition is, of course, passing the critical value $Q_c$).
SM Fig. 2. Timeseries of \( f \) for a finite-size system of \( N = 200, G = 23, Q = 24 \) (left); and \( G = 131, Q = 72 \) (right) observed in the model simulation on networks with \( K = 32 \). Here one Monte Carlo step consists of \( (N \times G) \times (N \times K)/2 \) opinion flips.

Details of the co-evolution based on the Hamiltonian in Eq. (4)

Starting from a random configuration of opinions and links, the society is updated from one timestep \( t \) to the next as follows:

1. Compute \( \bar{H} \) of the current state of the system, assume it has a value of \( \bar{H}_0 \).

2. Pick a node \( i \) at random and flip one of its opinions, \( a^i \). Compute \( \bar{H} \) again, it is now \( \bar{H}_1 \). If the value of \( \bar{H} \) has decreased in response to the flip, \( \bar{H}_1 \leq \bar{H}_0 \), accept the flip. If the value of \( \bar{H} \) increased, accept the flip only with probability \( p = e^{-\Delta \bar{H}} \), where \( \Delta \bar{H} = \bar{H}_1 - \bar{H}_0 \), if \( \bar{H}_1 > \bar{H}_0 \).

3. Compute \( \bar{H} \) of the system at this point, assume that it is now \( \bar{H}_0 \). We now pick one link randomly, \( J_{ij} \), and flip it. Compute \( \bar{H} \) again, and assuming it to be \( \bar{H}_1 \), we accept the flip if \( \bar{H}_1 \leq \bar{H}_0 \), and accept it with probability \( p' = e^{-\Delta \bar{H}} \), where \( \Delta \bar{H} = \bar{H}_1 - \bar{H}_0 \), if \( \bar{H}_1 > \bar{H}_0 \).

4. Continue with the next timestep.

Co-evolution based on the Hamiltonian in Eq. (4) with links determined by the opinions

One may ask what would happen to the co-evolutionary dynamics described in Eq. (4) if we update links after every opinion flip according to \( J_{ij} = \text{sign}(A_i \cdot A_j) \) in step 3. This means links are no longer random variables, but strictly determined by agent opinions. Still, those proposed changes that decrease the Hamiltonian in Eq. (4) are favoured over those that increase it, the same as before. Such a modified dynamics of Eq. (4) is rather similar to the presented model. Indeed, this variant can also be obtained if the same updating rule of the presented model with \( Q = N^\Delta \) and the stress in Eq. (4) are used instead of that in Eq. (1). SM Figure 3 shows the phase diagram of this variant of Eq. (4) (a) and its section at various values of \( g \) (b) obtained for a fully-connected network. A sharp transition between the cohesive and fragmented phases is also observed. The result suggests further research on a possible correspondence between the Hamiltonian and our agent-based approach if the effect of social balance is sufficiently large.

SM Fig. 3. Result of the model variant of Eq. (4) with strictly enforcing \( J_{ij} = \text{sign}(A_i \cdot A_j) \). (a) \( f \) is shown as a function of \( g \) and \( G \). (b) Section of the phase diagram for various strength of the coupling constant, \( g \). \( N = 100 \), results are averaged over 100 runs for fully-connected networks.