Continuous-time model of structural balance

Seth A. Marvel[a], Jon Kleinberg[b,1], Robert D. Kleinberg[b], and Steven H. Strogatz[a]

[a]Center for Applied Mathematics, and [b]Department of Computer Science, Cornell University, Ithaca, NY 14853

Edited by Ronald L. Graham, University of California, San Diego, La Jolla, CA, and approved November 30, 2010 (received for review September 3, 2010)

It is not uncommon for certain social networks to divide into two opposing camps in response to stress. This happens, for example, in networks of political parties during winner-takes-all elections, in networks of companies competing to establish technical standards, and in networks of nations faced with mounting threats of war. A simple model for these two-sided separations is the dynamical system $dX/dt = X^2$, where X is a matrix of the friendliness or unfriendliness between pairs of nodes in the network. Previous simulations suggested that only two types of behavior were possible for this system: Either all relationships become friendly or two hostile factions emerge. Here we prove that for generic initial conditions, these are indeed the only possible outcomes. Our analysis yields a closed-form expression for faction membership as a function of the initial conditions and implies that the initial amount of friendliness in large social networks (started from random initial conditions) determines whether they will end up in intractable conflict or global harmony.

random matrix theory | polarization

The mathematical model that we want to study is best understood as an outgrowth of a theory from social psychology known as *structural balance* (1). So let us begin with a brief explanation of what this theory says.

Consider three individuals: Anna, Bill, and Carl, and suppose that Bill and Carl are friends with Anna, but are unfriendly with each other. If the sentiment in the relationships is strong enough, Bill may try to strengthen his friendship with Anna by encouraging her to turn against Carl, and Carl might likewise try to convince Anna to terminate her friendship with Bill. Anna, for her own part, may try to bring Bill and Carl together so they can reconcile and become friends. In abstract terms, relationship triangles containing exactly two friendships are prone to transition to triangles with either one or three friendships.

Alternately, suppose that Anna, Bill, and Carl all view each other as rivals. In many such situations, there are incentives for the two people in the weakest rivalry to cooperate and form a working friendship or alliance against the third. In these cases, a single friendship may be prone to appear in a relationship triangle that initially has none.

These two thought experiments suggest a notion of stability, or balance, that can be traced back to the work of Heider (2). Heider's theory was expanded into a graph-theoretic framework by Cartwright and Harary (3), who considered graphs on n nodes (representing people, countries, or corporations) with edges signed either positive (+) to denote friendship or negative (−) to denote rivalry. If a social network feels the proper social stresses (those felt by Anna, Bill, and Carl in the examples above), then Cartwright and Harary's theory predicts that in steady state the triangles in the graph should contain an odd number of positive edges—in other words, three positive edges or one positive edge and two negative edges. We refer to such triangles as *balanced*, and triangles with an even number of positive edges as *unbalanced*. Finally, we call a graph *complete* if it contains edges between all pairs of nodes, and we say that a complete graph with signs on its edges is *balanced* if all its triangles are balanced. (All graphs in our discussion will be complete.)

As it turns out, these local notions of balance theory are closely related to the global structure of two opposing factions. In particular, suppose that the nodes of a complete graph are partitioned into two factions such that all edges inside each faction are positive and all edges between nodes in opposite factions are negative. (One of these factions may be empty, in which case the other faction includes all the nodes in the graph, and consequently all edges of the network are positive.) Note that this network must be balanced, because each triangle either has all three members in the same faction (yielding three positive edges) or has two members in one faction and the third member in the other faction (yielding one positive edge and two negative ones). In fact, a stronger and less obvious statement is true: Any balanced graph can be partitioned into two factions in this way, with one faction possibly empty (3). As a result, when we speak of balanced graphs, we can equivalently speak of networks with this type of two-faction structure.

Model

Structural balance is a static theory—it posits what a "stable" signing of a social network should look like. However, its underlying motivation is dynamic, based on how unbalanced triangles ought to resolve to balanced ones. This situation has led naturally to a search for a full dynamic theory of structural balance. Yet finding systems that reliably guide networks to balance has proved to be a challenge in itself.

A first exploration of this issue was conducted by Antal et al. (4) who considered a family of discrete-time models. In one of the main models of this family, an edge of the graph is examined in each time step, and its sign is flipped if this produces more balanced triangles than unbalanced ones. Although a balanced graph is a stable point for these discrete dynamics, it turns out that many unbalanced graphs called *jammed states* are as well (4, 5).

Thus, the natural problem became to identify and rigorously analyze a simple system that could progress to balanced graphs from generic initial configurations. An approach to this problem was taken by Kułakowski et al. (6), who proposed a continuous-time model for structural balance. They represented the state of a completely connected social network using a real symmetric $n \times n$ matrix X whose entry x_{ij} represents the strength of the friendliness or unfriendliness between nodes i and j (a positive value denotes a friendly relationship and a negative value an unfriendly one). Note that for a given X, there is a signed complete graph with edge signs equal to the signs of the corresponding elements x_{ij} in X. We will call X balanced if this associated signed complete graph is balanced.

Kułakowski et al. considered variations on the following basic differential equation, which they proposed as a dynamical system governing the evolution of the relationships over time:

Author contributions: S.A.M., J.K., R.D.K., and S.H.S. designed research; S.A.M., J.K., R.D.K., and S.H.S. performed research; S.A.M. analyzed data; and S.A.M., J.K., R.D.K., and S.H.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: kleinber@cs.cornell.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1013213108/-/DCSupplemental.

$$\frac{dX}{dt} = X^2. \quad [1]$$

Remarkably, simulations showed that for essentially any initial $X(0)$, the system reached a balanced pattern of edge signs in finite time.

Writing Eq. **1** directly in terms of the entries x_{ij} gives a sense of why this differential equation should promote balance:

$$\frac{dx_{ij}}{dt} = \sum_k x_{ik}x_{kj}. \quad [2]$$

Notice that x_{ij} is being pushed in a positive or negative direction based on the relationships that i and j have with k: If x_{ik} and x_{kj} have the same sign, their product guides the value of x_{ij} in the positive direction, whereas if x_{ik} and x_{kj} have opposite signs, their product guides the value of x_{ij} in the negative direction. In each case, this is the direction required to balance the triangle $\{i,j,k\}$. Note also that Eq. **2** applies for the case that $i = j$. Although this case is harder to interpret, the monotonic increase of x_{ii} implied by Eq. **2** might be viewed in psychological terms as an increase of self-approval or self-confidence as i becomes more resolute in its opinions about others in the network.

For a network with just three nodes, it can be easily proved that a variant of these dynamics generically balances the single triangle in this network; such a three-node analysis has been given by Kułakowski et al. (6), and we describe a short proof in *SI Text*. What is much less clear, however, is how the system should behave with a larger number of nodes, when the effects governing any one edge $\{i,j\}$ are summed over all nodes k to produce a single aggregate effect on x_{ij}.

It has therefore been an open problem to prove that Eq. **2** or any of the related systems studied by Kułakowski et al. will bring a generic initial matrix $X(0)$ to a balanced state. It has also been an open problem to characterize the structure of the balanced state that arises as a function of the starting state $X(0)$.

Results

In this paper, we resolve these two open problems. We first show that for a random initial matrix (with entries sampled independently from an absolutely continuous distribution with bounded support), the system reaches a balanced matrix in finite time with a probability converging to 1 in the number of nodes n. In addition, we provide a closed-form expression for this balanced matrix in terms of the initial one; essentially, we discover that the system of differential equations serves to "collapse" the starting matrix to a nearby rank-one matrix. We also characterize additional aspects of the process, giving, for example, a description of an "exceptional" set of matrices of probability measure converging to 0 in n for which the dynamics are not necessarily guaranteed to produce a balanced state.

We then analyze the solutions of the system for classes of random matrices in the large-n limit—in particular, we consider the case in which each unique matrix entry is drawn independently from a distribution with bounded support that is symmetric about a number μ (the mean value of the initial friendliness among the nodes). In this case, we find a transition in the solution as μ varies: When $\mu > 0$, the system evolves to an all-positive sign pattern, whereas when $\mu \leq 0$, the system evolves to a state in which the network is divided evenly into two all-positive cliques connected entirely by negative edges. We end by discussing some implications of the model and the associated transition between harmony and conflict, including an evaluation of the model on empirical data and some potential connections to research on reconciliation in social psychology.

Behavior of the Model: Evolution to a Balanced State

Suppose we randomly select the $x_{ij}(0)$'s from a continuous distribution on the real line. Then the $x_{ij}(t)$'s found by numerical integration generally sort themselves in finite time into the sign pattern of two feuding factions. To reformulate this observation as a precise statement and explain why the behavior holds so pervasively, we now solve Eq. **1** explicitly.

Solution to the Model. The initial matrix $X(0)$ is real and symmetric by assumption, so we can write it as $QD(0)Q^T$, where $D(0)$ is the diagonal matrix with the eigenvalues of $X(0)$, denoted $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, as diagonal entries ordered from largest to smallest, and Q is the orthogonal matrix with the corresponding eigenvectors of $X(0)$, denoted $\omega_1,\omega_2,\ldots,\omega_n$, as columns. The superscript T signifies transposition.

The differential equation Eq. **1** is a special case of a general family of equations known as *matrix Riccati equations* (7). The analysis of the full family is complicated and not fully resolved, but we now show that the special case of concern to us, Eq. **1**, has an explicit solution with a form that exposes its connections to structural balance. We proceed as follows. First, we observe that by separation of variables, the solution of the single-variable differential equation $\dot{x} = x^2$ (overdot representing differentiation by time) with initial condition $x(0) = \lambda_k$ is

$$\ell_k(t) = \frac{\lambda_k}{1 - \lambda_k t}. \quad [3]$$

Therefore the diagonal matrix $D(t) = \mathrm{diag}[\ell_1(t),\ell_2(t),\ldots,\ell_n(t)]$ is the solution of Eq. **1** for the initial condition $X(0) = \mathrm{diag}(\lambda_1, \lambda_2,\ldots,\lambda_n)$.

Moreover $Y(t) = QD(t)Q^T$ is also a solution of Eq. **1** because $\dot{Y} = Q\dot{D}Q^T = Q(D^2)Q^T = (QDQ^T)^2 = Y^2$. But $Y(t)$ has the same initial condition as $X(t)$ in our original problem: $Y(0) = QD(0)Q^T = X(0)$. So by uniqueness, $Y(t) = QD(t)Q^T$ must be the solution we seek.

Our solution $X(t)$ can also be written in a different way to mimic the solution of the one-dimensional equation $\dot{x} = x^2$. Because $x_{ij}(t) = \sum_{k=1}^n q_{ik}\ell_k(t)q_{jk}$, where q_{ij} is the (i,j)th entry of Q, we can expand the denominators of the $\ell_k(t)$ functions in powers of t to rewrite $X(t)$ as $X(0) + X(0)^2t + X(0)^3t^2 + \ldots$, or more concisely,

$$X(t) = X(0)[I - X(0)t]^{-1}. \quad [4]$$

(Note that the matrices $X(0)$ and $[I - X(0)t]^{-1}$ commute.) This equation is valid when t is less than the radius of convergence of every λ_k, that is, when $t < 1/\lambda_1$ (assuming $\lambda_1 > 0$).

Finally we note that the above method of solving Eq. **1** contains a reduction of the number of dynamical variables of the system from $\binom{n+1}{2}$ to n. The $\binom{n}{2}$ constants of motion generated by this reduction are just the off-diagonal elements of $Q^TX(t)Q = D(t)$, or $\sum_{k=1}^n \sum_{\ell=1}^n q_{ki}x_{k\ell}(t)q_{\ell j} = 0$ for all $1 \leq i < j \leq n$. Furthermore, the procedure for reducing $X(t)$ can be easily generalized to any system of the form $\dot{X} = f(X)$, where f is a polynomial of X.

Behavior of the Solution. Let us now examine the behavior of our solution $X(t)$ to see why in the typical case it splits into two factions in finite time. It turns out that this is the guaranteed outcome if the following three conditions hold (and as we see below, they hold with probability converging to 1 as n goes to infinity):

1. $\lambda_1 > 0$,
2. $\lambda_1 \neq \lambda_2$ (and hence $\lambda_1 > \lambda_2$), and
3. all components of ω_1 are nonzero.

To see why these conditions imply a split into two factions, observe from Eq. **3** that each $\ell_k(t)$ diverges to infinity at $t = 1/\lambda_k$. Because $x_{ij}(t) = \sum_{k=1}^n q_{ik}\ell_k(t)q_{jk}$, all x_{ij}'s diverge to infinity when

the ℓ_k with the smallest positive $1/\lambda_k$ does. Under the first and second conditions, this ℓ_k is ℓ_1, so the blow-up time t^* of Eq. **1** must be $1/\lambda_1$. To show that the nodes are partitioned into two factions as $X(t)$ approaches t^*, let $\overline{X}(t) = X(t)/||X(t)||$ on the half-open interval $[0,t^*)$, where $||X(t)||$ denotes the Frobenius norm of X. The matrix $\overline{X}(t)$ has the sign pattern of $X(t)$, and as t approaches t^* it converges to the rank-one matrix

$$X^* = Q\text{diag}(1,0,0,\ldots,0)Q^T = \omega_1\omega_1^T. \quad [5]$$

Now let ω_{1k} denote the value of the kth coordinate of ω_1, and let $S = \{k\colon \omega_{1k} > 0\}$ and $T = \{k\colon \omega_{1k} < 0\}$. Then S and T partition the node indices $1,2,\ldots,n$ by our condition that ω_1 has no zero components. From Eq. **5**, this partition must correspond to two cliques of friends joined by a complete bipartite graph of unfriendly ties.

The Three Conditions. We now return to the three conditions above. We first show that the second and third hold with probability 1. We then show that the first condition holds with probability converging to 1 as n goes to infinity. Finally, we analyze the behavior of the system in the unlikely event that the first condition does not hold. The fact that the conjunction of all three conditions holds with probability converging to 1 as n grows large justifies our earlier claim that the behavior described above holds for almost all choices of initial conditions.

First we show why the second and third conditions hold with probability 1 so long as the (joint) distribution from which $X(0)$ is drawn is absolutely continuous with respect to Lebesgue measure —in other words, assigns probability zero to any set of matrices whose Lebesgue measure is zero. Our arguments below make use of the following two basic facts:

i. The set of zeros of a nontrivial multivariate polynomial has Lebesgue measure zero, and
ii. the existence of a common root of two univariate polynomials P and Q is equivalent to the vanishing of a multivariate polynomial in the coefficients of P and Q (specifically, it is equivalent to the vanishing of the determinant of the Sylvester matrix of P and Q, also called the *resultant* of P and Q).

To show that $\lambda_1 \neq \lambda_2$ with probability 1, let P denote the characteristic polynomial of $X(0)$, and let Q denote the derivative of P. Then $X(0)$ has a repeated eigenvalue if and only if P has a repeated root, which it does if and only if P and Q have a common root. This condition is equivalent to the vanishing of the resultant of P and Q, which is a multivariate polynomial in the entries of $X(0)$. The polynomial cannot be zero everywhere, because there is at least one symmetric matrix that does not have a repeated eigenvalue. So the set of matrices having a repeated eigenvalue has Lebesgue measure zero.

Similarly, to show that all components of ω_1 are nonzero, let P denote the characteristic polynomial of $X(0)$ and P_i the characteristic polynomial of the $(n-1)\times(n-1)$ submatrix $X_i(0)$ obtained by deleting the ith row and ith column of $X(0)$. It is easy to check that if any eigenvector of $X(0)$ has a zero in its ith component, then the vector obtained by deleting that component is an eigenvector of $X_i(0)$ with the same eigenvalue. Consequently, P and P_i must have a common root, implying that the resultant of P and P_i vanishes. This resultant is once again a multivariate polynomial in the entries of $X(0)$, and once again it must be nonzero somewhere because there is at least one symmetric matrix whose eigenvectors all have nonzero entries. Hence, the set of matrices having an eigenvector with zero in its ith component has Lebesgue measure zero.

Finally, to determine the likelihood of the first condition, we first must say a bit more about the way that $X(0)$ is selected. Suppose that the off-diagonal $x_{ij}(0)$'s are drawn randomly from a common distribution F and the on-diagonal $x_{ii}(0)$'s are drawn randomly from a common distribution G. All selections are independent for $i \leq j$. [For $i > j$, we let $x_{ij}(0) = x_{ji}(0)$, so that $X(0)$ is symmetric.] For this construction of $X(0)$, Arnold (8) has shown that with the remarkably weak additional assumption that F has a finite second moment, Wigner's semicircle law holds in probability as n grows to infinity. This in turn implies that $\lambda_1 > 0$ in probability in the same limit.

Moreover, suppose we are in the low-probability case that $\lambda_1 \leq 0$. In this case, the analysis above shows that all the functions $\ell_i(t)$ converge to 0 as $t \to \infty$. Thus, $\lim_{t\to\infty} D(t) = 0$, and because $X(t) = QD(t)Q^T$, we also have $\lim_{t\to\infty} X(t) = 0$.

Although the entries of $X(t)$ converge to zero when $\lambda_1 \leq 0$, one might still want to know if the sign pattern of $X(t)$ is eventually constant (i.e., remains unchanged for all t above some threshold value) and, if so, what determines this sign pattern. It is possible to answer this question, again assuming the second and third conditions. By expanding the function $\ell_i(t) = \lambda_i/(1-\lambda_i t)$ in powers of $u = 1/t$, we obtain the asymptotic series

$$\ell_i(t) = -u - u^2\lambda_i^{-1} - O(u^3), \quad [6]$$

which implies

$$X(t) = QD(t)Q^T = -uI - u^2X(0)^{-1} - O(u^3). \quad [7]$$

In the limit of small u, the leading-order term of the diagonal entries of $X(t)$ is the linear term, which has negative sign. For the off-diagonal entries of $X(t)$, the leading-order term as u tends to zero is the quadratic term, whose sign matches the sign of the corresponding off-diagonal entry of the matrix $-X(0)^{-1}$.

Behavior of the Model: From Factions to Unification

The analysis in the previous section tells us how to find both the blow-up time t^* and final sign configuration of a network if we know its initial state $X(0)$. However, we might also want to know whether we can characterize the behavior of $X(t)$ in the large-n limit in terms of statistical parameters of $X(0)$. This could, for example, help us forecast the behavior of groups of individuals when collecting complete relationship-level data is not feasible. Clearly if the underlying network is a complete graph, it is not realistic to consider n that are too large, but we find fortunately in simulations that the asymptotic behavior we will derive in this section becomes apparent even for moderate values of n (less than 100). As a result, these large-n results are perhaps most usefully viewed as an approximate guide to what happens in medium-sized groups that are large enough to show predictable collective behavior but for which a completely connected set of relationships is still feasible to maintain.

In this section, we show that there is a transition from final states consisting of two factions to final states consisting of all-positive relations as the "mean friendliness" of $X(0)$ (the mean of the distributions used to generate the off-diagonal entries of $X(0)$) is increased from negative to positive values. This is consistent with the numerical simulations shown in Fig. 1; as noted above, the asymptotic behavior we are studying is already clear in these simulations, which are performed with $n = 90$.

Before discussing the details, we describe how $X(0)$ is selected in this section. We start by adopting the procedure of Füredi and Komlós (9): The elements $x_{ij}(0)$ are drawn independently from distributions F_{ij} with zero mass outside of $[-K,K]$. The off-diagonal F_{ij}'s have a common expectation μ and finite variance σ^2, whereas the on-diagonal F_{ii}'s have a common expectation ν and variance τ^2. In addition, we require that each off-diagonal distribution F_{ij} be symmetric about μ. Random matrix models of this type have attracted considerable recent interest (see, e.g., refs. 10 and 11), but we need only the basic results of Füredi and Komlós (9) for our purposes, and so we use these in the

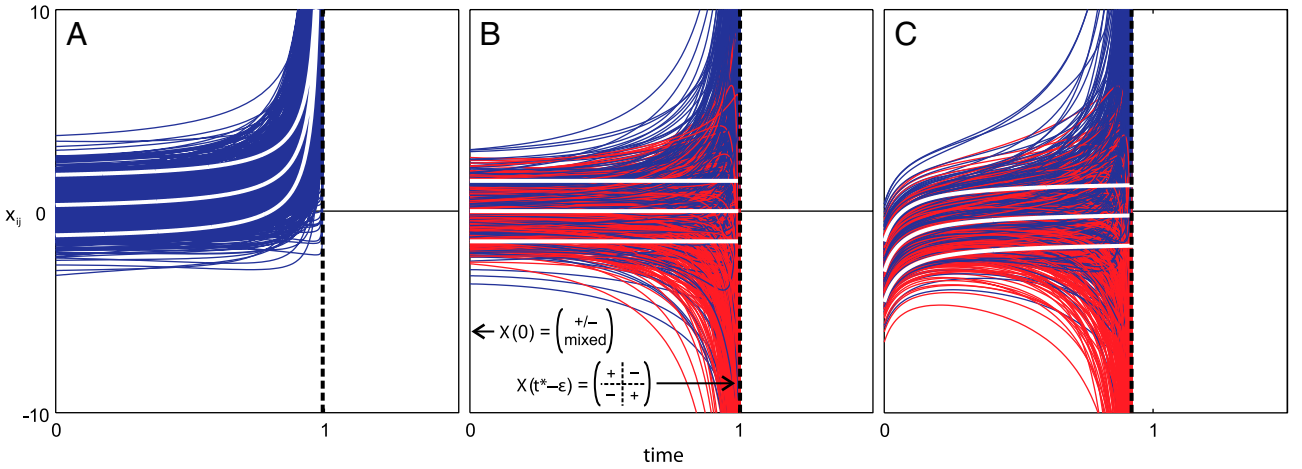


Fig. 1. Representative large-n plots of the model for (A) $\mu > 0$ ($\mu = 3/10$ in the plot shown), (B) $\mu = 0$, and (C) $\mu < 0$ ($\mu = -3$ in the plot shown). For all three plots, $\sigma = 1$ and $n = 90$. To reduce image complexity, only one randomly sampled fifth of the trajectories is included. In the second plot, t^* denotes the time at which the system diverges, and ε denotes a sufficiently small displacement. The white curves superimposed on the three plots are the large-n trajectories $x_{ij}(t) = x_{ij}(0) - \mu + \mu/(1 - \mu nct)$ for $x_{ij}(0) = \mu, \mu \pm 3\sigma/2$, where c represents a rescaling of time. Because we want to fix the blow-up time t^* near 1 and because $ct^* = 1/\lambda_1$ as found in the text, we choose $c = 1/(\mu n + \nu - \mu + \sigma^2/\mu)$ for A and $c = 1/(2\sigma\sqrt{n})$ for B and C using estimates of λ_1 taken from ref. 9. The black dotted lines mark the blow-up times $t^* = 1/(c\lambda_1)$.

development below. We consider the three cases of positive, zero, and negative μ.

Case 1: $\mu > 0$. The results of Füredi and Komlós (9) show that when $\mu > 0$, the deviation of ω_1 from $(1,1,\ldots,1)/\sqrt{n}$ vanishes in probability in the large-n limit. Hence the final state of the system consists of one large clique of friends containing all but at most a vanishing fraction of the nodes. Moreover, by assuming a bound on σ we can strengthen this statement further: If $\sigma < \mu/2$, then the findings of Füredi and Komlós imply that the final state consists of a single clique of friends, with no negative edges. These observations are consistent with the representative numerical trial shown in Fig. 1*A*. Moreover, Füredi and Komlós show that the asymptotic behavior of λ_1 grows like $\mu n + O(1)$, and hence the blow-up time scales like $1/(\mu n)$.

We can gain insight into the behavior of the system for small t using an informal Taylor series calculation: If we rescale time in Eq. **1** by inserting a $1/n$ before the summation, compute the Taylor expansion of $x_{ij}(t)$ term by term, and then take the expectation of each term, we obtain the geometric series $x(t) = \mu + \mu^2 t + \mu^3 t^2 + \cdots$, or

$$x(t) = \frac{\mu}{1 - \mu t}. \qquad [8]$$

With significantly more work, it can be proved that every trajectory $x_{ij}(t)$ has this time dependence on $[0,1/K)$ in the large-n limit with probability 1 (see *SI Text*), so we may write

$$\lim_{n\to\infty} x_{ij}(t) = x_{ij}(0) - \mu + \frac{\mu}{1 - \mu t} \quad \text{with prob. 1} \qquad [9]$$

for all t in $[0,1/K)$. Observe that this limit has a blow-up time t^* of $1/\mu$. Because our rescaling of time represents a zooming in or magnification of time by a factor of n, this t^* corresponds to a blow-up time asymptotic to $1/(\mu n)$ for the unrescaled system, consistent with the results of Füredi and Komlós.

Case 2: $\mu = 0$. In the event that the network starts from a mean friendliness of zero, numerical experiments indicate that the system ends up with two factions of equal size in the large-n limit (Fig. 1*B*). We now prove this to be the case. For the remainder of this discussion, we abbreviate $X(0)$ as A and $x_{ij}(0)$ as a_{ij}.

Because the off-diagonal entries of A have symmetric distributions by assumption, we have for any off-diagonal a_{ij} and any interval S_{ij} on the real line that $P(a_{ij} \in S_{ij}) = P(-a_{ij} \in S_{ij})$. Now let D be a diagonal matrix with some sequence of $+1$ and -1 along its diagonal (where the ith diagonal entry is denoted by d_i). Then the random matrices A and $B = DAD$ are identically distributed, as we now show.

To say that A and B are identically distributed means that for every Borel set of matrices S, $P(A \in S) = P(B \in S)$. To prove this, it suffices to consider the case in which S is a product of intervals S_{ij}, because these product sets generate the Borel σ-algebra. The upper triangular entries of A are independent, so $P(A \in S) = \prod_{i\le j} P(a_{ij} \in S_{ij})$. Similarly, $P(B \in S) = \prod_{i\le j} P(d_i a_{ij} d_j \in S_{ij})$. By the symmetry of the off-diagonal distributions, $\prod_{i\le j} P(a_{ij} \in S_{ij}) = \prod_{i\le j} P(d_i a_{ij} d_j \in S_{ij})$, which gives us $P(A \in S) = P(B \in S)$ as desired. (Note that when $i = j$, the factor $d_i d_j$ is 1 so the on-diagonal distributions need not be symmetric.)

Now consider the set S of matrices with an ω_1 consisting of all-positive components. The above demonstration implies that the probability of choosing an A in this set is the same as choosing an A such that B is in this set. Regarding the later event, $A(D\omega_i) = \lambda_i(D\omega_i)$ implies $B\omega_i = \lambda_i\omega_i$, so the λ_1 eigenvector of the A used to compute B is $D\omega_1$. This demonstrates that all sign patterns for the components of ω_1 are equally likely. In other words, the distribution of the number of positive components in ω_1 is the binomial distribution $B(n,1/2)$ and the fraction of positive components in ω_1 converges (in several senses) to $1/2$ as n grows large.

Additionally, we can consider how λ_1 varies with n in the case that $\mu = 0$ to determine when the blowup will occur. Füredi and Komlós (9) found for this case that $\lambda_1 \in 2\sigma\sqrt{n} + O(n^{1/3}\log n)$ with probability tending to 1, so with probability tending to 1 the blow-up time shrinks to zero like $1/\sqrt{n}$, an order of $\sqrt{n}$ slower than in the $\mu > 0$ case.

Case 3: $\mu < 0$. For this final case, Füredi and Komlós (9) found that $\lambda_1 < 2\sigma\sqrt{n} + O(n^{1/3}\log n)$ with probability tending to 1. The semicircle law gives a lower bound: $\lambda_1 > 2\sigma\sqrt{n} + o(\sqrt{n})$ in probability. So the blow-up time goes to zero like $1/\sqrt{n}$ in the unrescaled system.

Note also that if we define a matrix $C = -A$, where A is now the initial matrix $X(0)$ of Case 3, then C satisfies the condition of

Case 1, $\mu > 0$. Thus the distance between the top eigenvector of C and $(1,1,\ldots,1)/\sqrt{n}$ declines to zero in probability just as in Case 1. Furthermore, every other eigenvector of C is orthogonal to the largest one. Hence if $\sigma < |\mu|/2$, then with probability tending to 1, every other eigenvector acquires a mixture of positive and negative components in the large-n limit, including the bottom eigenvector of C, which is the top eigenvector of A. This establishes that in the case that $\mu < 0$ and $\sigma < |\mu|/2$, the system ends up in a state with two factions with probability converging to 1 for all finite n.

Numerical simulations of the case that $\mu < 0$ suggest the conjecture that the two factions are approximately equal in size for large n. Furthermore, the derivation of Eq. **9** is in fact valid for all μ, so each trajectory rapidly decays from $x_{ij}(0)$ toward $x_{ij}(0) - \mu$ on $[0,1/K)$ (Fig. 1*C*). This transient decay appears to extend beyond $t = 1/K$ in numerical simulations. So, for example, if time is rescaled by $1/\sqrt{n}$ instead of $1/n$, we would hypothesize that (*i*) each trajectory makes a complete jump from $x_{ij}(0)$ to $x_{ij}(0) - \mu$ in the large-n limit and that (*ii*) from this point onward, the system behaves like an initial configuration of the $\mu = 0$ case and so separates into two equal factions *en route* to its blowup at $1/(2\sigma)$.

Discussion

In this final section, we review our results and their significance relative to previous work in structural balance theory. We then compare the predictions of the model with data, discuss potential criticisms of the model, and finish with some intriguing connections between the behavior of the model and recent social-psychological work on neutralizing two-sided conflicts.

Our first result is a demonstration that the model forms two factions in finite time across a broad set of initial conditions. As noted at the outset, similar demonstrations have not been possible for dynamic models of structural balance in earlier literature because these models contained so-called *jammed states* that could trap a social network before it reached a two-faction configuration (4, 5). The model of Kułakowski et al. by contrast has no such jammed states for generic initial conditions and hence provides a robust means for a social network to balance itself.

The second result of the paper is the discovery and characterization of a transition from global polarization to global harmony as the initial mean friendliness of the network crosses from nonpositive to positive values. Similar transitions have been observed in other models of structural balance but so far none has been characterized at a quantitative level. For example, Antal et al. (4) found a nonlinear transition from two cliques of equal size to a single unified clique as the fraction of positively signed edges at $t = 0$ was increased from 0 to 1 (see figure 5 of ref. 4). The authors provided a qualitative argument for this transition, but left open the problem of its quantitative detail. Our results both confirm the generality of their observations and provide a quantitative account of a transition analogous to theirs.

To complement the theoretical nature of our work and get a better sense of how the model behaves in practice, we can numerically integrate it for several cases of empirical social network data where the real-life outcomes of the time evolution are known. Our first example is based on a study by Zachary (12) who witnessed the breakup of a karate club into two smaller clubs. Prior to the separation, Zachary collected counts of the number of social contexts in which each pair of individuals interacted outside of the karate club, with the idea being that the more social contexts they shared, the greater the likelihood for information exchange. These counts, or *capacities* as Zachary called them, can be converted to estimates of friendliness and rivalry in many different ways. For a large class of such conversions, Eq. **1** predicts the same division that Zachary's method found, which misclassified only 1 of the 34 club members (Fig. 2 *A* and *B*).

A second example can be constructed from the data of a study by Axelrod and Bennett (13) regarding the aggregation of Allied and Axis powers during World War II. If we simply take the entries of their propensity$(i,j) \cdot$ size$(i) \cdot$ size(j) matrix to be proportional to the friendliness felt between the various pairs of countries in the war, then running the model gives the correct Allied-Axis split for all countries except Denmark and Portugal (Fig. 2*C*).

Nevertheless, the model clearly contains several strong simplifications of the underlying social processes. The first of these is inherent to structural balance theory itself; it is a framework restricted to capture a particular kind of social situation, in which the need for consistency among one's friendships and rivalries brings about the emergence of two factions. Extensions of the theory have considered models in which it is possible to have multiple mutual enemies and hence more than two factions (14), and also networks that are not complete graphs (3). However, our focus here has been on the basic theory, because as we have seen, obtaining a satisfactory dynamics even for this simplest form of structural balance has been an elusive challenge. Moreover, the

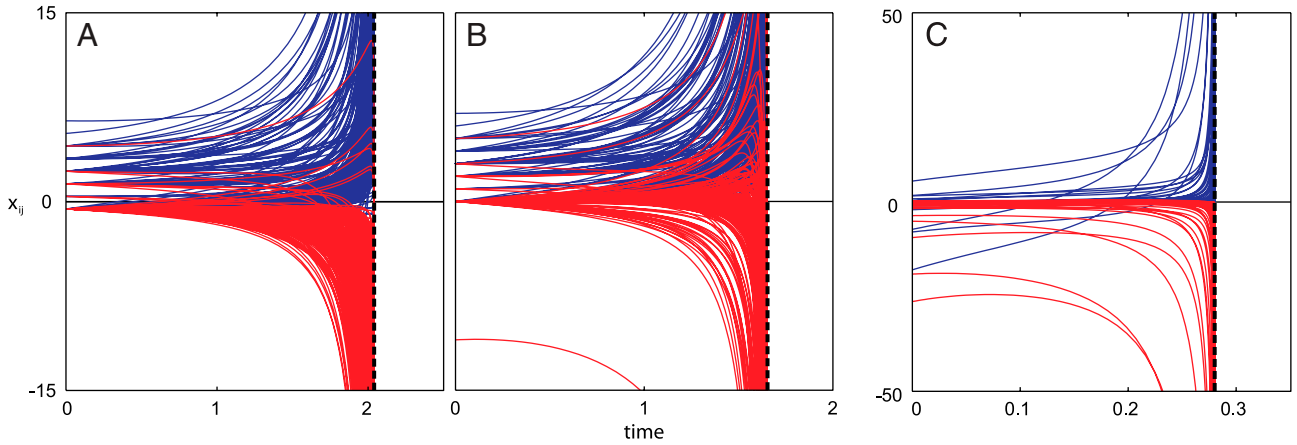


Fig. 2. Tests of the model of Kułakowski et al. (Eq. **1**) against two existing datasets. (*A*) The evolution of the model starting from Zachary's capacity matrix with the capacity of each relationship reduced by 0.58. This is the minimal downward displacement necessary (to two significant figures) for the resulting separation to be correct for all but 1 of the 34 club members. For reasons described by Zachary (12), this is basically the best separation we can expect. (*B*) The evolution of the model from Zachary's capacity matrix with the capacity of zero between the two club leaders replaced by −11; the resulting factions are identical to those in *A*. Substituted values less than −11 yield the same two factions, whereas greater values produce less accurate divisions. (*C*) The evolution of the model starting from Axelrod and Bennett's 1939 propensity$(i,j) \cdot$ size$(i) \cdot$ size(j) matrix for the 17 countries involved in World War II (by Axelrod and Bennett's definition). The model finds the correct split into Allied and Axis powers with the exceptions of Denmark and Portugal. Axelrod and Bennett's own landscape theory of aggregation does slightly better—its only misclassification is Portugal.

basic version of structural balance that we have considered here, with a complete graph of relationships and constraints leading to two factions, is relevant to a range of different situations. These span the kinds of settings discussed earlier in this section, including clubs, classrooms, and small organizations (15), as well as international relations during crisis (where a large set of nations can all mutually maintain friendly or unfriendly diplomatic relations) (13, 16).

Another consequence of the particular model studied here that has no direct analogue in real social situations is the divergence to infinity of the relationship strengths x_{ij}. However, because the purpose of the model is to study the pattern of signs that emerges, our main conclusions are based not on the actual magnitudes of these numbers but on the fact that the sign pattern eventually stabilizes at a point before the divergence. This stabilization of the sign pattern is our primary focus, and one could interpret the subsequent singularity as simply the straightforward and unimpeded "ramping up" of values caused by the system once all inconsistencies have been worked out of the social relations—the divergence itself can be viewed as taking place beyond the window of time over which the system corresponds to anything real. Alternately, one can imagine that as the community completes its separation into two groups, other social processes take over. For example, individuals with differing ideological views or social preferences may self-segregate, breaking the all-to-all assumption of the model. In other cases, mounting tensions may erupt into violence, reflecting a sort of bound on the relationship intensity achievable for pairs of nodes in the network.

Finally, we note that there is a large body of work in social psychology that studies issues such as the formation and reconciliation of factions from a much more empirical basis; see, for example, refs. 17 and 18. It is an interesting open problem to determine the extent to which the strictly mathematical development of the models here can be combined with the perspectives in this empirical body of literature, ultimately leading to a richer theory of these types of social processes.

ACKNOWLEDGMENTS. We thank Nick Trefethen and David Bindel for pointers to the literature on matrix Riccati equations. Research was supported in part by the John D. and Catherine T. MacArthur Foundation, a Google Research Grant, a Yahoo! Research Alliance Grant, an Alfred P. Sloan Foundation Fellowship, a Microsoft Research New Faculty Fellowship, a grant from the Air Force Office of Scientific Research, and National Science Foundation Grants CCF-0325453, BCS-0537606, CCF-0643934, IIS-0705774, and CISE-0835706.

1. Wasserman S, Faust K (1994) *Social Network Analysis: Methods and Applications, Structural Analysis in the Social Sciences* (Cambridge Univ Press, New York), pp 220–248.
2. Heider F (1946) Attitudes and cognitive organization. *J Psychol* 21:107–112.
3. Cartwright D, Harary F (1956) Structural balance: A generalization of Heider's theory. *Psychol Rev* 63:277–293.
4. Antal T, Krapivsky PL, Redner S (2005) Dynamics of social balance on networks. *Phys Rev E* 72:036121.
5. Marvel SA, Strogatz SH, Kleinberg JM (2009) Energy landscape of social balance. *Phys Rev Lett* 103:198701.
6. Kułakowski K, Gawroński P, Gronek P (2005) The Heider balance—a continuous approach. *Int J Mod Phys C* 16:707–716.
7. Abou-Kandil H, Freiling G, Ionescu V, Jank G (2003) *Matrix Riccati Equations in Control and Systems Theory* (Birkhäuser, Basel, Switzerland) p 21.
8. Arnold L (1971) On Wigner's semicircle law for the eigenvalues of random matrices. *Probab Theory Rel* 19:191.
9. Füredi Z, Komlós J (1981) The eigenvalues of random symmetric matrices. *Combinatorica* 1:233–241.
10. Vu V (2007) Spectral norm of random matrices. *Combinatorica* 27:721–736.
11. Tao T, Vu V (2010) Smooth analysis of the condition number and the least singular value. *Math Comput* 79:2333–2352.
12. Zachary WW (1977) An information flow model for conflict and fission in small groups. *J Anthropol Res* 33:452–473.
13. Axelrod R, Bennett DS (1993) A landscape theory of aggregation. *Brit J Polit Sci* 23:211–233.
14. Davis JA (1967) Clustering and structural balance in graphs. *Hum Relat* 20(2):181–187.
15. Doreian P, Mrvar A (1996) A partitioning approach to structural balance. *Soc Networks* 18:149–168.
16. Moore M (1978) An international application of Heider's balance theory. *Eur J Soc Psychol* 8:401–405.
17. Pettigrew TF (1998) Intergroup contact theory. *Annu Rev Psychol* 49:65–85.
18. Pettigrew TF, Tropp LR (2006) A meta-analytic test of intergroup contact theory. *J Pers Soc Psychol* 90:751–783.