

Critical properties of Heider balance on multiplex networks

Krishnadas Mohandas ^{*}, Krzysztof Suchecki [†], and Janusz A. Hołyst [‡]
Faculty of Physics, Warsaw University of Technology, Koszykowa 75, 00-662 Warsaw, Poland



(Received 10 October 2023; accepted 29 February 2024; published 9 April 2024)

Heider's structural balance theory has proven invaluable in comprehending the dynamics of social groups characterized by both friendly and hostile relationships. Since people's relations are rarely single faceted, we investigate Heider balance dynamics on a multiplex network, consisting of several copies of the same agent displaying correlated relations at different layers building the multiplex. Intralayer interactions in our model adhere to Heider dynamics, while interlayer correlations stem from Ising interactions, with the heat-bath dynamics of link signs. Our investigation reveals a multifaceted system with a diverse equilibrium landscape contingent on the coexistence of distinct phases across layers. We observe that, starting from a paradise state with positive links in all layers, an increase in temperature triggers a discontinuous transition to a disordered state akin to single-layer scenarios. The critical temperature surpasses that of the single-layer case, a fact verified through extended mean-field analysis and agent-based simulations. Furthermore, the scenario shifts when one layer exhibits a two-clique configuration instead of a paradise state. This change introduces additional transitions: synchronization of interlayer relations and a transition to the disorder, appearing at a different, lower temperature compared to matching paradise states. This exploration shows the intricate interplay of Heider balance and multiplex interactions.

DOI: [10.1103/PhysRevE.109.044306](https://doi.org/10.1103/PhysRevE.109.044306)

I. INTRODUCTION

Multiplex structures [1] find widespread application in characterizing diverse systems such as social groups, transportation networks, and biological frameworks, including protein-protein interaction systems [2–8]. Humans are inherently complex beings, and when they form interconnected groups bound by relationships, the intricacy only deepens. Individuals frequently maintain a plethora of relationships, varying in nature from familial and professional ties to friendships and online connections. A multiplex network representation is invaluable in addressing this diversity of relations [4,9,10]. Through distinct layers or edge types, this representation delineates the assorted interaction types. Incorporating the multiplex essence of these relationships provides an enriched comprehension of the system's dynamics. In certain instances, this approach unveils hitherto undiscovered mechanisms or phenomena that had remained concealed within simple, aggregated network models [11,12].

When examining human interactions, a fundamental categorization involves distinguishing between friendly and hostile relations. Within a group of individuals, these relations can be effectively portrayed using signed networks [13]. Positive links denote friendly relations between nodes representing individuals, while negative links signify hostile relations. In this context, Heider introduced the structural balance theory (SBT) to social psychology as a means to

delineate the underlying tensions within such networks [14]. Central to Heider's theorem is the principle that a signed network attains balance when all triads exhibit either three positive relations or one positive and two negative relations [15,16].

Balance theory posits that networks characterized by friend-or-foe relationships evolve towards a state of greater equilibrium [17]. Notably, Cartwright and Harary demonstrated that, beyond an all-positive paradise state, a complete graph is balanced if agents can be segregated into two groups, featuring solely positive ties within each group and exclusively negative ties bridging the groups [18,19].

Recent studies have extended the concept of structural balance to the realm of statistical physics, establishing a parallel between the tension of imbalanced triads and energy excitation above the system's ground state [20–22]. This framework enables the incorporation of uncertainty regarding individual actions as thermal noise, quantifiable through temperature. Adopting this approach obviates the need for specific microscopic dynamical rules governing changes in relations. Furthermore, it harnesses established statistical methods from physics to describe and predict system behavior.

Several models have been developed, grounded in the principles of SBT, aiming to elucidate human behavior within social networks [23–27]. However, the exploration of structural balance within the realm of multiplex networks remains limited to a handful of investigations [28–32].

Within this study, we extend the Heider balance (HB) concept to the domain of multiplex networks, wherein nodes (agents) are linked by various kinds of relations corresponding to different network layers. Imagine a case of a group of people that work for the same company and, simultaneously,

^{*}krishnadas.mohandas.dokt@pw.edu.pl

[†]krzysztof.suchecki@pw.edu.pl

[‡]janusz.holyst@pw.edu.pl

privately belong to a sports club where they also know each other. The social ties between people in different contexts (work or club) correspond to links in different layers, with the network nodes being the same across layers and representing the same people. Each layer of the signed network describes relations in a specific context, and there are no signed links going across layers. It is possible for a relation between people to depend on context: Two people may be competing “enemies” within the company while at the same time being friends on their own time while interacting in the club. However, our premise is rooted in the idea that individuals exhibit a predilection for the consistency of various relation types. If they share a friendly link within a work context, this amiability is likely to extend beyond the workplace and vice versa. There is an influence of relation in one layer on the relation in the other, and we assume these influences are symmetric and can be represented as a tension between incoherent relations in different contexts. To capture this concept, we introduce a coupling mechanism that fosters similarity among relations linking the same pair of individuals, akin to the coupling between spins in the Ising model of ferromagnets.

Our objective is to understand the impact of layer-to-layer coupling on the signed relation configurations across diverse layers. We start by considering a duplex network configuration, where intralayer relations in each layer form a complete graph and interlayer Ising interactions are exclusive to the same link replicas. Subsequently, we generalize this model to encompass an arbitrary number of layers. The dynamics of relation signs within each layer is governed by Heider structural balance dynamics. Conversely, Ising dynamics dictates the interlayer coupling of relation signs concerning the same pairs of individuals.

The forthcoming section extends the classical mean-field theory of HB to the context of multiplex networks. Our theoretical approach correctly predicts the occurrence of a discontinuous order-disorder transition induced by thermal noise, including both duplex and multiplex network cases. The subsequent section employs Monte Carlo simulations to validate the analytical findings. Further investigation includes a more thorough overview of HB dynamics within a duplex network, where layers can exist in different states not described by the analytical approach. The dynamics of such a system encompasses a temperature-driven transition toward synchronization between layers in different states and a subsequent transition toward disorder.

II. MODEL

Our model represents people as agents placed in vertices of a multiplex network, with dynamical relations between them represented by signed edges. Edges belonging to a specific layer correspond to some type or context of relation (such as workplace or private relations), while the same vertices exist across all layers. The vertices representing agents have no dynamical attributes, while edge signs obey Heider structural balance related to interactions within each layer and Ising-type dynamics related to interactions between layers to take into account positive correlations between different relations for the same pair of agents.

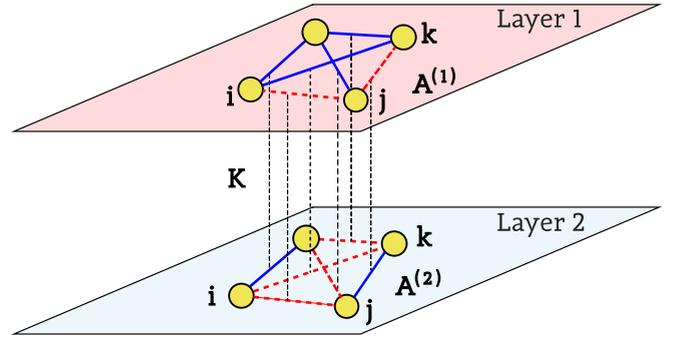


FIG. 1. Example of interactions between edge signs (blue solid line for +1 and red dashed line for -1) on a duplex network. Edge sign changes are driven by intralayer interactions according to Heider structural balance (with strength $A^{(1)}$ or $A^{(2)}$, respectively) as well as interlayer coupling of Ising nature between the edges, with strength K . The signs of relations evolve over time, according to the heat-bath dynamics with energy as expressed by Eq. (1) and given temperature T .

Let us consider a multiplex network with N nodes and L layers. For simplicity, we assume that every pair of agents has a relation between them, meaning that the topology of each layer is a complete graph. This situation corresponds to a small social system where everybody knows everybody. Each pair of agents i and j has therefore L signed relations $x_{ij}^{(\alpha)} = \pm 1$ between them, where i and j indicate agents and $\alpha = 1, 2, \dots, L$ indicates the layer. The edge signs $x_{ij}^{(\alpha)}$ are dynamic variables and can change over time. Let us stress that the nodes representing people are the same across all layers; the layers represent different relation contexts between the same people, not different groups of people. We assume that each imbalanced triad (ij, jk, ki) of relations causes a certain tension that the agents i , j , and k involved try to relax. This is represented as a specific energy associated with the triad: $-A^{(\alpha)}$ if the triad is balanced and $+A^{(\alpha)}$ if it is not, according to Heider structural balance theory. Energy $A^{(\alpha)} > 0$ represents the strength of the Heider coupling between link signs within each layer. Similarly, a discrepancy between relations in different layers between the same pair of agents i and j causes a tension, also represented as energy: $-K$ if the relations are the same and $+K$ if they are different. Energy $K > 0$ corresponds to the ferromagnetic Ising model coupling between link signs across layers. An important detail to note is that the coupling described by K constant is between link states, not nodes, and is therefore not expressed as links. Figure 1 illustrates the model for a duplex network.

The Hamiltonian H of the system, or its energy E , is

$$H = - \sum_{\alpha=1}^L A^{(\alpha)} \sum_{i>j>k} x_{ij}^{(\alpha)} x_{jk}^{(\alpha)} x_{ki}^{(\alpha)} - K \sum_{\alpha>\beta} \sum_{i>j} x_{ij}^{(\alpha)} x_{ij}^{(\beta)}. \quad (1)$$

The state of the system will change according to the Hamiltonian in the presence of temperature T , which represents the uncertainty of relations or tolerance towards imbalanced relations in the social systems modeled.

The model has degenerate ground states, where any division of vertices into two groups is a ground state, if all intragroup edges in all layers are positive and all intergroup edges are negative [15]. We will call this state a two-clique state. This also includes the so-called paradise state, where one group contains all vertices and the other no vertices so that all relations are positive. Due to thermal fluctuations, at any temperature $T > 0$ the system settles into an equilibrium different from the ground state and in fact is multistable, with the exact state reached via a dynamical process depending not only on parameters but also on the initial conditions.

III. ANALYTIC APPROACH

A. Mean-field approximation for a duplex network

Let us consider a duplex network, a particular case of a system described by the Hamiltonian (1) corresponding to $L = 2$ where each of the two link layers has the structure of the complete graph. A mean-field approach for a single network was proposed in [33], where every single link x_{ij} interacts with the mean-field proportional to $\langle x \rangle^2$ as a result of the presence of this link in all $N - 2$ triads. Here angular brackets indicate an ensemble average over all possible states. This method works accurately for a single-layer complete graph with $N \gg 1$ since each individual link interacts with a significant fraction of all other links, which is accurately captured by the mean-field approach. Extending this approach to multiplex networks is not as straightforward as it may seem at first glance. A naive inclusion of the second layer's influence as an Ising mean field would represent each edge having an Ising-type interaction with all the edges of the second layer, instead of only the single edge it actually should interact with. This can drastically alter the behavior of the system, especially for strong interlayer interactions.

In order to tackle this for a duplex network, a mean-field approach is employed with a pair of coupled links $\bar{x}_{ij} = [x_{ij}^{(1)}, x_{ij}^{(2)}]$ instead of a single link $x_{ij}^{(1)}$ as our elementary subsystem that will be interacting with the mean field. Our pair \bar{x}_{ij} will be experiencing interlayer interaction as internal energy of the state \bar{x}_{ij} and intralayer interactions (in both layers) as the interaction of \bar{x}_{ij} with a two-dimensional mean field proportional to $[\langle x^{(1)} \rangle^2, \langle x^{(2)} \rangle^2]$. Since we assume the

system dynamics is expressed by state energy and thermal equilibrium in a specific temperature T , the probabilities of states \bar{x}_{ij} of this pair are described by the canonical ensemble

$$P(\bar{x}_{ij}) = \frac{\exp[-E(\bar{x}_{ij})/T]}{\sum_{\bar{x}_{mn}} \exp[-E(\bar{x}_{mn})/T]}, \quad (2)$$

where

$$E(\bar{x}_{ij}) = -A^{(1)} \sum_{k \neq i, j} x_{ij}^{(1)} x_{jk}^{(1)} x_{ik}^{(1)} - A^{(2)} \sum_{k \neq i, j} x_{ij}^{(2)} x_{jk}^{(2)} x_{ik}^{(2)} - K x_{ij}^{(1)} x_{ij}^{(2)}. \quad (3)$$

We will assume the same interaction parameters for all the layers, i.e., $A^{(1)} = A^{(2)} = A$. The expected value of $\langle \bar{x}_{ij} \rangle$ can be written as

$$\langle \bar{x}_{ij} \rangle = \sum_{\bar{x}_{ij}} P(\bar{x}_{ij}) \bar{x}_{ij}. \quad (4)$$

Following the mean-field method, we assume that the link sign $x_{ij}^{(\alpha)}$ interacts not with a specific link sign product $x_{jk}^{(\alpha)} x_{ki}^{(\alpha)}$ but with a mean field $(x^{(\alpha)})^2$, with the mean link signs $x^{(\alpha)}$ that will be defined as

$$\bar{x} \equiv [x^{(1)}, x^{(2)}]. \quad (5)$$

We call this mean link sign value a polarization of the network (if considering the entire vector) or of the given layer (if considering specific components $x^{(\alpha)}$ of it). This allows us to write the energy of a given state \bar{x}_{ij} as

$$E(\bar{x}_{ij}) = -A(N - 2)x_{ij}^{(1)}(x^{(1)})^2 - A(N - 2)x_{ij}^{(2)}(x^{(2)})^2 - Kx_{ij}^{(1)}x_{ij}^{(2)}. \quad (6)$$

Combining (6) and (2) into (4) allows us to calculate the expected value of link signs $\langle \bar{x}_{ij} \rangle$ in the considered pair ij (see Appendix). Assuming that the calculated $\langle \bar{x}_{ij} \rangle$ is the same as the \bar{x} in the mean field, meaning $\langle \bar{x}_{ij} \rangle = \bar{x}$, allows us to write a set of self-consistent equations for mean polarization \bar{x} ,

$$\begin{aligned} x^{(1)} &= f_1(x^{(1)}, x^{(2)}), \\ x^{(2)} &= f_2(x^{(1)}, x^{(2)}), \end{aligned} \quad (7)$$

where

$$f_\alpha(x^{(1)}, x^{(2)}) = \frac{e^{2d} \sinh \{a[(x^{(1)})^2 + (x^{(2)})^2]\} + \sinh \{a(-1)^\alpha [(x^{(2)})^2 - (x^{(1)})^2]\}}{e^{2d} \cosh \{a[(x^{(1)})^2 + (x^{(2)})^2]\} + \cosh \{a[(x^{(1)})^2 - (x^{(2)})^2]\}}, \quad (8)$$

with $\alpha \in \{1, 2\}$. The parameters $a = \frac{AM}{T}$ and $d = \frac{K}{AM}$ are rescaled intralayer and interlayer interaction strengths with $M = N - 2$ the number of triads each edge belongs to. Figure 2 shows the numerical solution of average link polarization based on the mean-field solution [solution of Eq. (7)] for different values of rescaled interlayer coupling strength d . We observe from Fig. 2 that by increasing temperature T , the mean polarization continuously decreases to a point x_c when it jumps to zero. At this point, a first-order transition is observed at a critical temperature T_c . The values T_c and x_c depend on the coupling strength K between the layers. Increasing the

coupling strength increases T_c , asymptotically approaching a certain saturated T_c value (as seen in Fig. 5). Note that the critical temperature T_c represents the temperature where a stable polarized state ($\bar{x} \neq 0$) disappears. Since the unpolarized state is always stable in the mean-field approach, this first-order transition is always from a polarized to an unpolarized state, the reverse transition does not take place, and an unpolarized system will remain unpolarized even if the temperature is reduced below T_c . The actual behavior of the system, outside the mean-field approach, is more complex (see Sec. V).

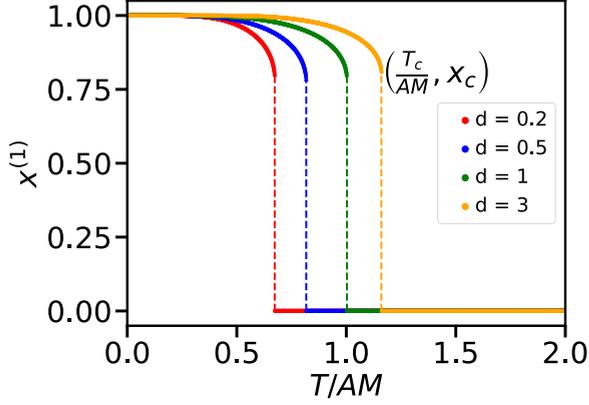


FIG. 2. Mean-field solution for average link polarization $x^{(1)}$ for various interlayer coupling strengths. The coupled layers undergo a discontinuous transition at a critical temperature T_c that increases with coupling strength. The results are shown for $d = 0.2, 0.5, 1, 3$ with the transition point moving from left to right with increasing d .

To tackle the problem analytically, one can treat Eqs. (7) as recurrence equations describing how the average link polarizations $[x^{(1)}, x^{(2)}]$ evolve when influenced by the mean field and its past state. This allows us to consider the issue of multistability that our system exhibits and how it could change between these states, instead of being limited to just finding equilibria. The dynamical equations are

$$\begin{aligned} x^{(1)}(t+1) &= f_1(x^{(1)}(t), x^{(2)}(t)), \\ x^{(2)}(t+1) &= f_2(x^{(1)}(t), x^{(2)}(t)), \end{aligned} \quad (9)$$

where $f_\alpha(x^{(1)}(t), x^{(2)}(t))$ are the right-hand sides of Eqs. (7) taken at time t . Then the fixed points of the map [Eqs. (9)] are solutions of the implicit equations (7). It is worth noting that $\bar{x}_0 = [0, 0]$ is always a stable solution of Eqs. (9), which for lower temperatures can coexist with another stable solution. When considering the behavior of the system, we are looking at the critical temperature T_c where the nonzero solution disappears and the stable state of the system starting in paradise discontinuously changes from x_c to 0.

The values T_c and x_c can be obtained from a pair of transcendental algebraic relations (10) and (11) that describe the fixed point and its Jacobian matrix at the point where the largest eigenvalue of the Jacobian crosses 1 (the critical point). Detailed calculations are shown in Appendix. The relation between T_c and x_c is given by the auxiliary variable $z = e^{(AM/T_c)x_c^2}$ that obeys

$$8 \ln z = \frac{(z^4 - 1)(z^4 D^2 + 2z^2 + D^2)}{z^2(z^4 + 2z^2 D^2 + 1)}, \quad (10)$$

where $D = e^d$. Solving this equation numerically allows us to find z satisfying it, which depends only on D and thus indirectly on the ratio of interlayer coupling K and intralayer coupling A (Fig. 15 in the Appendix shows the general shape of this solution).

Given the z value, it is possible to find critical polarization $x_c(D, z)$ and the temperature $T_c(x_c, z)$,

$$\frac{T_c}{AM} = \frac{x_c^2}{\ln z} = \left(\frac{1}{\ln z} \right) \left(\frac{D^2(z^4 - 1)}{D^2(z^4 + 1) + 2z^2} \right)^2. \quad (11)$$

The dependence of this critical temperature on coupling can be seen in Fig. 5. For $D = 1$ (noninteracting layers), the approach is effectively reduced to the one-layer case and Eq. (11) gives the critical temperature of a single-layer network.

B. Generalization to a higher-order multiplex network

In order to study multilayer effects, we generalize the duplex network considered in Sec. III A in the following way. We consider L layers of the same set of nodes and assume that each layer corresponds to a different type of relationship or communication context. Edges in each layer interact with corresponding edges in all other layers since different relationships of the same pair of agents are correlated. For simplicity, the same parameters $A^{(\alpha)} = A$ are used for all the layers and the coupling strength between each pair of layers is the same K . The energy of all L interactions between a given pair of agents ij is

$$E(\bar{x}_{ij}) = -A \left(\sum_k \sum_{\alpha=1}^L x_{ij}^{(\alpha)} x_{jk}^{(\alpha)} x_{ki}^{(\alpha)} \right) - K \sum_{\alpha > \beta} x_{ij}^{(\alpha)} x_{ij}^{(\beta)}. \quad (12)$$

The approach used to analyze duplex networks can be used for a higher number of layers. For $L = 3$, using Eq. (4) with $\bar{x}_{ij} = (x_{ij}^{(1)}, x_{ij}^{(2)}, x_{ij}^{(3)})$ and energy (12) for the three-layer network and following the same methodology as for duplex network, we can write a set of self-consistent equations for $\bar{x} = (x^{(1)}, x^{(2)}, x^{(3)})$. Using a method analogous to that for a duplex network (see the Appendix), we arrive at the transcendental equation for an auxiliary variable $z = z(D) = e^{(AM/T_c)x_c^2}$,

$$8 \ln(z) = \frac{(D^4 z^6 + z^4 - z^2 - D^4)(D^4 z^6 + 3z^4 + 3z^2 + D^4)}{z^2(D^4 z^8 + 4D^4 z^6 + 3D^8 z^4 + 3z^4 + 4D^4 z^2 + D^4)}, \quad (13)$$

with the critical temperature for $L = 3$ being

$$\frac{T_c}{AM} = \frac{1}{\ln z} \left(\frac{D^4(z^6 - 1) + z^2(z^2 - 1)}{D^4(z^6 + 1) + 3z^2(z^2 + 1)} \right)^2. \quad (14)$$

The dependence of this critical temperature on coupling can be seen in Fig. 5 and on the number of layers in Fig. 6. While this approach can be used for any specific L , considering L -dimensional \bar{x} , the equations quickly become intractable with increasing L .

To make a general prediction for the critical temperature T_c where the order in the system disappears, the states of all layers are assumed to be statistically the same. This allows us to simplify our approach. Instead of considering the exact microstate \bar{x}_{ij} of the set of links between i and j , we consider a mesostate described only by the number of positive links L^+ among all L links in the set. Since by assumption all layers are statistically the same, the exact placement of positive and negative signs within the L copies does not matter, and the energy of a microstate depends only on the mesostate variable L^+ , so the dynamics of the system can be described entirely through that variable. For a set of L links, with L^+ positive and L^- negative links, then $L^- = L - L^+$ and the mean polarization

is

$$\langle x \rangle = \left\langle \frac{L^+ - L^-}{L} \right\rangle = \frac{2\langle L^+ \rangle}{L} - 1, \quad (15)$$

where $\langle x \rangle$ is the average polarization of links in each of the layers $\langle x \rangle = \langle x_{ij}^{(\alpha)} \rangle$, with the mean over all links and ensemble, and α being any of the layers (they are statistically identical by assumption). Because the number of microstates aggregated into a given mesostate L^+ is not fixed but is equal to the number of ways in which L^+ positive links are distributed across L total, we need to consider this number of microstates contributing to a given mesostate L^+ when calculating canonical ensemble probabilities. The number of microstates \mathcal{M} for mesostate L^+ is given by

$$\mathcal{M} = \binom{L}{L^+}. \quad (16)$$

The energy of state L^+ for a set of links between a given node pair is the sum of the Heider energy of links interacting with different layers as well as the Ising energy of interaction between different links in the set. The Heider energy is according to Eq. (12), but instead of specific $x_{ij}^{(\alpha)}$ terms we have L^+ positive and L^- negative components in the sum over layers and instead of all $x_{jk}^{(\alpha)}$ and $x_{ki}^{(\alpha)}$ terms we have mean-field polarization x ,

$$E_{\text{Heider}}(L^+) = -A(N-2)[L^+(x)^2 - (L-L^+)(x)^2]. \quad (17)$$

One should be cautious not to mistake the macroscopic mean-field variable $\langle L^+ \rangle$ that appears in the expression for $\langle x \rangle$ with the microscopic state variable L^+ . The distinction can be dropped only when calculating the mean value of microscopic L^+ later in Eq. (20).

The Ising energy corresponds to the sum over all pairs of links, L^+ of which are positive and L^- negative, giving

$$E_{\text{Ising}}(L^+) = -K \frac{L^+(L^+ - 1) + (L - L^+)(L - L^+ - 1)}{2} + KL^+(L - L^+), \quad (18)$$

with the complete energy being the sum of these two

$$E(L^+) = E_{\text{Heider}}(L^+) + E_{\text{Ising}}(L^+). \quad (19)$$

Hence the mean number of positive links can be calculated from the canonical ensemble, multiplying the factor for specific L^+ by the number of microstates such a state actually represents. At this point, the mean value of microscopic L^+ is considered the same as the mean-field variable $\langle L^+ \rangle$, which means we obtain a self-consistent equation for $\langle L^+ \rangle$,

$$\langle L^+ \rangle = g(\langle L^+ \rangle) = \frac{\sum_{L^+=0}^L L^+ \binom{L}{L^+} e^{-E(L^+)/T}}{\sum_{L^+=0}^L \binom{L}{L^+} e^{-E(L^+)/T}}. \quad (20)$$

Here $\langle L^+ \rangle = L$ is a paradise state and $\langle L^+ \rangle = L/2$ is a completely disordered state with half the links positive and half negative. Due to using a mesostate description, instead of L equations we have a simple scalar equation (20) that could be solved using a similar stability analysis approach. Figure 3 shows a graphical representation of the right-hand side of Eq. (20), whose solution is given by the intersection

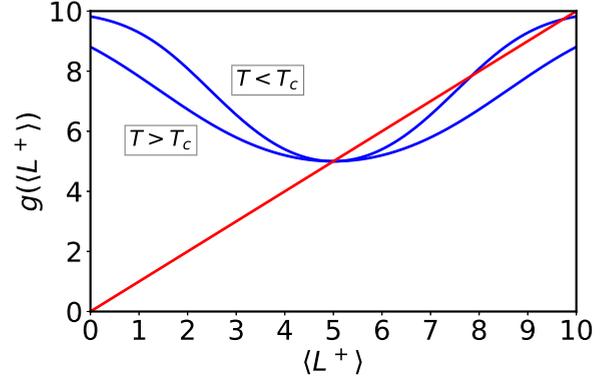


FIG. 3. Graphical solution of the mean-field state equation (20) for $L = 10$ layers. The blue solid curve is $y = g(\langle L^+ \rangle)$ and the red straight line is $y = \langle L^+ \rangle$. When $T > T_c$ there is a single stable fixed point. When $T < T_c$ there are three fixed points and the inner one is unstable

of the straight line L^+ and right-hand side. For $T > T_c$ there is only one stable fixed point $L^+ = L/2$, which corresponds to a disordered state of the system, while for $T < T_c$ an additional stable fixed point appears, corresponding to an ordered state, as well as an unstable fixed point separating basins of attraction for each stable solution. When $T = T_c$ both additional points are the same unstable fixed point tangential to the diagonal line. The critical temperature can be numerically estimated from the fixed-point equation (20) and its derivative $dg(\langle L^+ \rangle)/d\langle L^+ \rangle = 1$. The L coupled layers undergo a first-order transition at T_c for any number of layers, which is confirmed numerically for up to ten layers using Eq. (20).

C. Saturation of critical temperature

The critical temperature $\frac{T_c}{AM}$ depends on d (relative coupling strength) and it increases with an increase of this coupling strength, saturating for high d values (see Fig. 5).

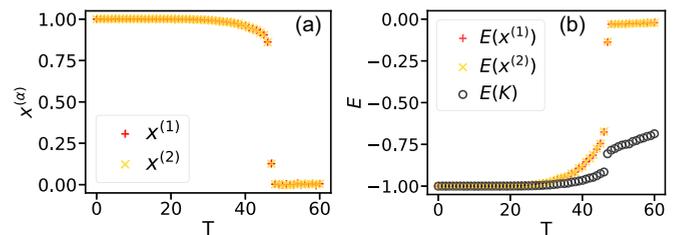


FIG. 4. Mean link polarizations $x^{(1)} = \langle x_{ij}^{(1)} \rangle$ and $x^{(2)} = \langle x_{ij}^{(2)} \rangle$ (a) and corresponding energies (b) for a duplex network as a function of temperature T for the complete graphs with $N = 50$ nodes and a coupling strength $K = 50$ (corresponding to $d = 1$). The results show averages over 50 independent simulations. Initial conditions were paradise states at both layers, where the overlapping red pluses and yellow crosses correspond to layers 1 and 2, respectively. In contrast to the intralayer energies (overlapping red pluses and yellow crosses), the interlayer energy (black circles) increases slowly when $T > T_c$, with layers remaining partially synchronized even when intralayer disorder sets in.

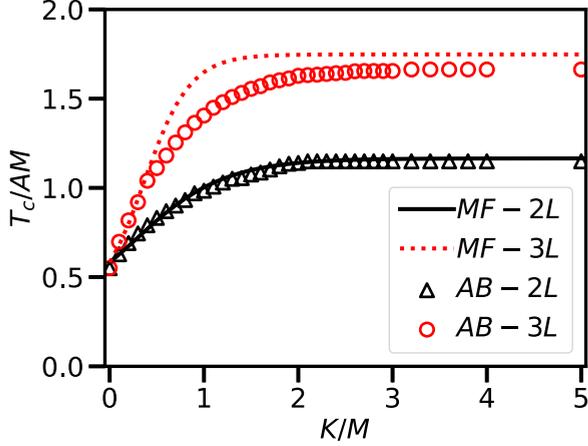


FIG. 5. Critical temperature T_c increases with coupling strength K and saturates to $T \approx LT_{c,L=1}$ for high K/M values. The black solid line and black triangles show the mean-field prediction (11) and numerical results for $L = 2$, respectively, while the red dotted line and red circles show mean-field prediction (14) and numerical results for $L = 3$, respectively.

For large coupling strength $D \rightarrow +\infty$, Eqs. (10) and (11) can be reduced to

$$8 \ln z = \frac{z^8 - 1}{2z^4} \Rightarrow z \approx 1.72346 \dots, \quad (21)$$

$$\frac{T_c}{AM} = \frac{1}{\ln z} \left(\frac{z^4 - 1}{z^4 + 1} \right)^2 \Rightarrow \frac{T_c}{AM} \approx 1.16516 \dots \quad (22)$$

Comparing this to the results for a single layer, where $T_c/AM \approx 0.58258$ (in agreement with [33]), we find that

$$T_c|_{L=2, d \rightarrow +\infty} \approx 2T_c|_{L=1}. \quad (23)$$

To ensure the generality of the observed relationship (23) between the number of layers and the critical temperature, our study extends to a system with three layers $L = 3$. In the limit $D \rightarrow +\infty$, Eqs. (13) and (14) can be reduced to

$$8 \ln z = \frac{z^{12} - 1}{3z^6} \Rightarrow z \approx 1.43749 \dots, \quad (24)$$

$$\frac{T_c}{AM} = \frac{1}{\ln z} \left(\frac{z^6 - 1}{z^6 + 1} \right)^2 \Rightarrow \frac{T_c}{AM} \approx 1.74516 \dots \quad (25)$$

The obtained saturation critical temperature satisfies the relation

$$T_c|_{\alpha=3, d \rightarrow +\infty} \approx 3T_c|_{\alpha=1}. \quad (26)$$

For L coupled layers, the transcendental equation at $D \rightarrow +\infty$ could be used to obtain the saturation critical temperature

$$8 \ln z = \frac{z^{4L} - 1}{Lz^{2L}}, \quad (27)$$

$$\frac{T_c}{AM} = \frac{1}{\ln z} \left(\frac{z^{2L} - 1}{z^{2L} + 1} \right). \quad (28)$$

Equation (28). allows us to conclude that

$$T_c|_{\alpha=L, K \rightarrow +\infty} \approx LT_c|_{\alpha=1}. \quad (29)$$

Thus the saturation critical temperature increases proportionally to the number of layers L .

IV. NUMERICAL SIMULATIONS OF PARADISE STATE STABILITY

For the verification of the analytical calculations presented in the preceding section, numerical Monte Carlo simulations of the model are performed. We use a heat-bath algorithm, with each update changing an entire set of links between randomly selected agents i and j and updates happening asynchronously, with $N(N-1)/2$ elementary updates treated as a time step. Each update consists of randomly choosing a pair of agents ij with uniform probability and then changing the link signs of all parallel links \vec{x}_{ij} to one of 2^L possible states, with probabilities corresponding to the canonical ensemble probabilities $P(\vec{x}_{ij}) = 1/Z \exp[-E(\vec{x}_{ij}/k_B T)]$. The energy $E(\vec{x}_{ij})$ is calculated using the current state of the rest of the entire system (meaning all \vec{x}_{kl} where $kl \neq ij$), according to Eq. (12). The initial condition is always started with a fully connected graph in the paradise configuration (all links positive). The heat-bath algorithm is used rather than the more well-known Metropolis algorithm due to the extremely slow convergence of the single-link update Metropolis algorithm for a larger number of layers and relatively strong interlayer coupling. It is worth noting that for $L = 2$ the results of simulations using both algorithms are the same, up to the fluctuations inherent in the stochastic Monte Carlo methods.

For the complete graph, the number of pair neighbors of nodes ij is equal to $M = N - 2$ and we keep $A^{(1)} = A^{(2)} = 1$. The intralayer energy is rescaled to the average value of Heider energy per triad whereas the interlayer energy is scaled to the average energy value per interacting link pair.

The simulations are repeated for a range of temperatures from $T = 0$ with step ΔT to find the highest value of T for which $\langle x_{ij}^{(\alpha)} \rangle$ is positive. The true value of T_c lies in the interval $[T^*, T^* + \Delta T]$ and the estimated value of critical temperature is $T_c = T^* + \Delta T/2$.

For the initial condition of the paradise state, the system moves to a disordered state above temperature T_c . Figure 4 shows that both layers behave in the same aligned way. The intralayer energies $E(x^{(1)})$ and $E(x^{(2)})$ approach zero above the critical temperature and the mean polarization values fluctuate around zero. The interlayer energy $E(K)$ is negative both above and below the critical temperature, showing that the link signs in different layers are partially aligned even in the disordered state above T_c . The critical temperature T_c predicted from numerically evaluated analytical mean-field theory is compared to those obtained from Monte Carlo simulations (Fig. 5). In spite of the slight shift of the transition when comparing numerical simulations with that of the mean-field approximation, the analytical method is found to be successful. Figure 6 shows how the critical temperature varies with the number of layers where the simulations also validate Eq. (29) in the saturation limit. Here there is some discrepancy between our predictions and simulation results for a large number of layers. The analytical approach predicts that as L increases, even if coupling K is small, at some point the critical temperature will start rising nonlinearly, increasing all the way up to the same saturated critical temperature T_c , regardless of K . Even low K values show this behavior for a sufficiently large number of layers L . This behavior can be best seen in Fig. 6 for $K = 10$ and 20. Simulations show,

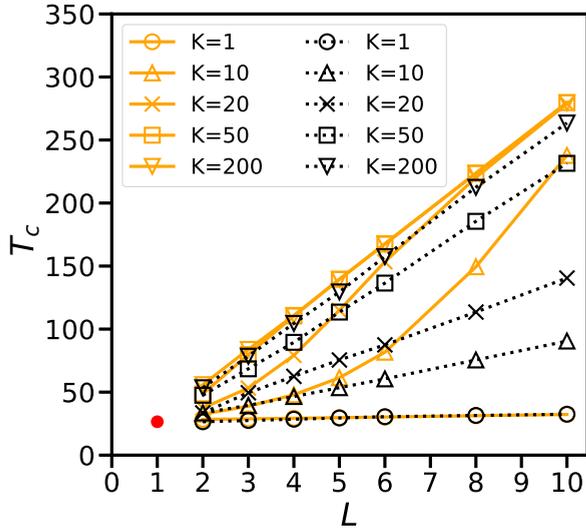


FIG. 6. Critical temperature T_c increases with the number of layers L when coupling K is positive. Black symbols and dotted lines show the results of Monte Carlo simulations, while orange symbols and solid lines show the mean-field prediction [iterating Eq. (20) as a map] at different interlayer coupling strengths for $N = 50$ and $A = 1$. The critical temperature for a single layer is shown as a closed red circle at $T_c = 26.5$.

however, that no such transition exists and the critical temperature increases linearly with L for any K , with a slope depending on K . This means that our mean-field approach is unable to fully and accurately predict the behavior of the system with many weakly interacting layers, although it is still qualitatively correct. Our predictions work for both low K and L , where the analytical approach predicts a linear increase of T_c with L , as well as high K values, where the temperature is essentially always saturated.

V. CRITICAL BEHAVIOR OF THE SYSTEM WITH LAYERS IN DIFFERENT STATES

The critical temperature investigated so far was defined as the temperature where the nonzero stable state disappears. The analytical approaches only describe the existence of a paradise state and look into its stability and transition into disorder. The multistability of the system means that in the range of parameters where paradise is stable, other states may exist, and starting from different initial conditions, the system may settle into different stable states. In this section we explore what happens in a multiplex network in more detail, focusing on situations where the two layers are in different configurations.

For the Heider balance dynamics, there are many specific states that are fully balanced, but it can be described in general as a two-clique state, where vertices are split into two groups, with all internal links in these groups positive and all links between groups negative. The specific states can differ by the relative sizes of the groups, from evenly sized up to a specific state where one group consists of all vertices and the other has none, which is the paradise state. The division of the entire network into two cliques of approximately the same size is the most likely to emerge from disorder out of all the possible

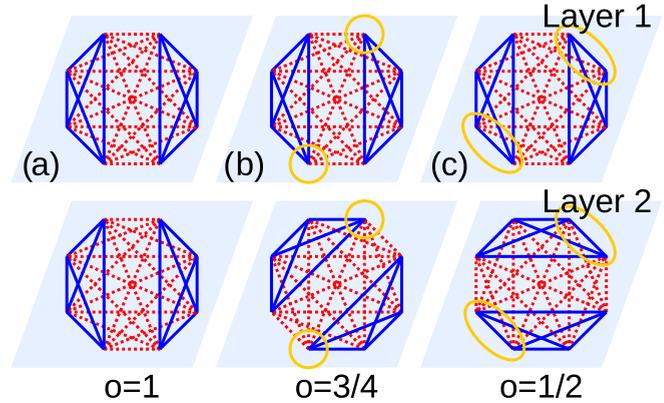


FIG. 7. If both layers of a duplex are in two-clique states, the overlap between the layers can vary from (a) full overlap (with matching order and overlap $o = 1$) through (b) partial overlap (with partial matching order and overlap $o = 3/4$ in this example) to (c) minimal overlap displayed by orthogonal two-clique states (with mismatched order and overlap $o = 1/2$). The same nodes exist in both layers and interlayer interactions exist only between the same link replicas with coupling strength K . Blue solid lines represent positive links, red dotted lines represent negative links, and yellow circles enclose agents belonging to different groups in both layers.

balanced states as a consequence of Heider balance dynamics [15] and thus is the most likely balanced state to be present. The paradise is extremely unlikely to emerge but represents a notable state, where every agent is friendly with another, similar to consensus in opinion dynamics. We limit our investigation to consider only three possible equilibrium states, namely, the two-clique state (with groups of equal sizes), the paradise state, and the disordered state, and ignore states with unequal group sizes.

In the case of a duplex network, the possible states that are balanced according to intralayer Heider balance dynamics (ignoring unequal-size two-clique states) consist of both layers being in the paradise state, one being a paradise and the other being a two-clique state, or both layers being in a two-clique state. Both layers in the paradise states have matching link signs across layers since all of them are positive. The state of one layer featuring paradise and the other two cliques has mismatched link signs across layers: Approximately half of the links (between groups) in the two-clique state are negative, while the corresponding links in the layer in paradise are positive, resulting in a net interlayer energy close to zero. The dual two-clique state is a special case, in the sense that the grouping of agents into cliques in both layers can display a varying degree of correlation. Two two-clique states can range from both layers having the same groupings (matching order), through vertex groupings that are partially correlated (partial matching order), down to minimal overlap between groupings (mismatched order), as shown in Fig. 7. We measure how different the two-clique states in two layers are by overlap o , which is a fraction of vertices belonging to the same groups in both layer or, in other words, how much the assignments of vertices to cliques in both layers overlap. Note that, due to groups themselves having no identity, we always consider groups that have higher overlap to be the same, so the overlap o is limited to the range $[\frac{1}{2}, 1]$. This overlap is directly tied to the

Spearman correlation between groups an agent belongs to in both layers (this correlation is equal to $c = 2o - 1$), as well as to the interlayer energy of fully Heider balanced states (neglecting fluctuations) $E_{\text{inter}} = \frac{KN}{2}[1 - N(1 - 2o)^2]$, where N is the number of agents and K is the interlayer coupling strength introduced in Sec. II. Note that, due to the discrete nature of agents, only some specific discrete overlap values are possible for a given specific N , but they will all be contained within the range $[\frac{1}{2}, 1]$. Overall, a duplex system in equilibrium at a specific temperature can be in a state of matching order with both layers in the paradise state or matching two-clique states with $o = 1$, a state of partial matching order with two-clique states of varying overlap $o \in (\frac{1}{2}, 1)$, a completely mismatched order state of two-clique states with overlap $o = \frac{1}{2}$ or two-clique and paradise layers, and finally a disordered state where both layers are disordered instead of being internally balanced. In the case of a multiplex with a higher number of layers, there are more states, with each pair featuring a different degree of overlap, such as a matching pair of layers and a third layer that does not match, three mutually mismatched layers, or a combination of partial overlaps. Note that there are certain restrictions on overlaps, such as a perfect match between layers being transitive; it is, for example, impossible for layers 1 and 2 to match or for layers 2 and 3 to match while 1 and 3 do not match, but there may be many different mutually mismatched layers, depending on a total number of agents N . For the easiest case of $N = 2^n$ agents, there can be as many as $1 + \binom{N/2}{N/4}$ two-clique states with overlap $o = 0$ with each other.

The analytical approaches introduced in Sec. III A do not consider the existence of two-clique states (other than the special case, the paradise). The conclusions from the in-depth numerical investigation presented here may differ significantly from mean-field predictions whenever two-clique states are involved. The most important difference is that the mean field predicts a disordered state to be always stable, even at low temperatures, which is not true in reality, as is shown later in this section. In simulations, the system orders into a Heider balanced two-clique state under such conditions, which also possesses zero polarization, just like the disordered state. Because the analytical approach does not take the existence of two-clique states into account, it is unable to predict any critical behavior involving two-clique states that is presented in this section. In particular, the temperatures T_s , T_d , T_o , and T_m introduced later in the section are outside the scope of our analytical approach.

We consider and investigate numerically five scenarios where the initial condition contains mismatched layers: (i) One layer is in the paradise state while the other is a two-clique balanced state, (ii) both layers are in a mutually mismatched two-clique balanced state, (iii) both layers are in two-clique states that are partially matching, (iv) a three-layer network has one layer in the paradise state and two layers in matching two-clique states, and (v) a three-layer network has one layer in the paradise state and two other layers in mismatched two-clique states. Based on the results obtained in these situations as well for the initially disordered system, we draw conclusions regarding what system states exist in what range of parameters, what are transitions between them, and the overall shape of the phase diagram for the system.

A. Coupling between a paradise and a two-clique state

Consider a mismatched order state containing paradise in one layer and a two-clique state in the second. Assume that the size of the entire group is an even number $N = 2m$ and that each hostile group has the size m . Each node possesses $m - 1$ positive links to agents in its own clique and m negative links to the second clique. The behavior of the system depends on temperature and interlayer interaction strength, with four out of the total of five characteristic or critical temperatures visible from the simulation results shown in Fig. 8.

The temperature T_s is the temperature above which the mismatched order loses stability. In consequence, above T_s the initially mismatched layers will synchronize with each other, creating a matching order state. Above T_o it is possible that mismatched layers will become disordered instead of creating a matching order state. For higher interaction strength K this temperature coincides with T_d [Fig. 8(c)], which is a minimum temperature required for a disordered state to be stable. If the interaction strength is low, as $K \rightarrow 0$ these temperatures diverge, with T_o going towards single-layer critical temperature T_c (where order loses stability) and T_d going towards single-layer critical temperature T_d (where disorder gains stability), as shown on Fig. 13. Note that above T_o , whether the system ends up in a matching order state or a disordered state is essentially random, with probabilities depending on temperature T and layer overlap. The temperature T_m is an upper bound for two mismatched layers to evolve towards a matching order state. Above T_m the clash of two mismatched layers always results in a disordered state. Note that $T_m < T_c$, so the initial mismatched order evolves to disorder at a lower temperature than the initial matching order, which can persist up to temperature T_c (the only temperature not seen in Fig. 8). We consider T_s and T_d as critical temperatures, since certain states gain or lose stability, while at T_o and T_m no such thing happens. Hence we do not refer to T_m and T_o as critical temperatures but as characteristic temperatures.

The outcome of synchronization of the initially mismatched order state is not always the same and contains randomness. In our case of a mismatched paradise and two-clique state, the outcome could be either a matching paradise or a matching two-clique state. In addition, the resulting two-clique state does not need to match the state of the initial two-clique layer; the cliques could be of different sizes and placement than initially, which suggests both layers may undergo reconfiguration, instead of one layer imposing its own state on the other. The mean link polarization of a two-clique state depends on the sizes of cliques m_1 and m_2 ,

$$\langle x_{ij} \rangle = \frac{\binom{m_1}{2} + \binom{m_2}{2} - m_1 m_2}{\binom{N}{2}}, \quad (30)$$

where $\langle x_{ij} \rangle$ would be near 1 for clique sizes $m_1 \gg m_2$ and close to 0 for $m_1 \approx m_2$. We quantitatively investigated the frequency f of mean link polarization for both network layers after the synchronization event. Figure 9 shows the distribution of mean link polarization. Even though the paradise state $\langle x_{ij} \rangle = 1$ is the most likely result of the synchronization, two-clique states are also probable results.

We have performed simulations for different interlayer coupling strengths K , allowing us to determine how critical

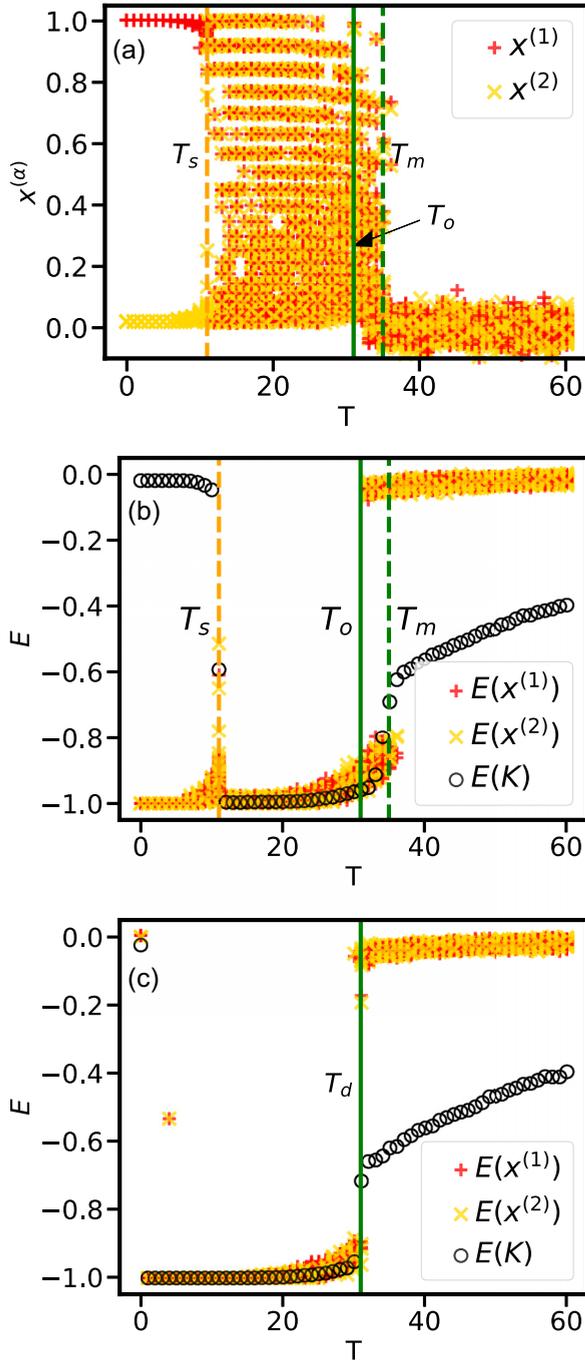


FIG. 8. In the case of coupled paradise and two-clique states, the system shows two transitions: synchronization of layers starting at T_s and transition to disorder starting at T_o and ending at additional characteristic temperature T_m . (a) Mean link polarization of 50 independent simulations over a range of temperatures for $N = 50$ nodes and a coupling strength $K = 25$, with one layer starting in the paradise state (red pluses) and the second layer starting in the two-clique state (yellow crosses). (b) Corresponding mean intralayer energy (red pluses and yellow crosses) and interlayer energy (black circles). (c) Intralayer and interlayer energies for a duplex network starting in a disordered state, showing that spontaneous order appears at any temperature below T_d .

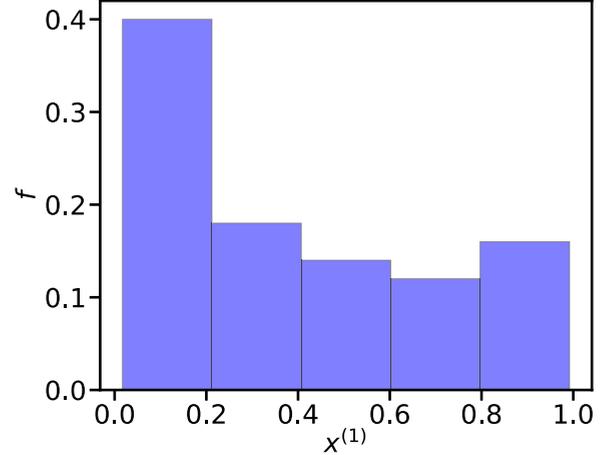


FIG. 9. When paradise and two-clique states synchronize, it results most often in a paradise state $x^{(1)} \approx 1$ but two-clique states $x^{(1)} < 1$ are also a possible outcome. The graph shows the distribution of mean link polarization after paradise and two-clique states synchronize at the temperature T where $T_s < T < T_d$.

temperatures T_d and T_s as well as characteristic temperatures T_o and T_m depend on model parameters. These results have been aggregated into the phase diagram for two-layered system, found in Sec. V D.

Note that when the system results in a disordered state, it is still partially synchronized due to the interlayer Ising interactions. This is equivalent to paramagnetic partial ordering due to the external field, which in this case is the impact of the other layer. In effect, the interlayer energy $E(K)$ in Fig. 8 only asymptotically approaches zero with increasing temperature, unlike intralayer energy related to internal layer ordering.

An interesting observation during synchronization is how the link polarization of the two layers evolves over time. Figure 10 displays examples of time trajectories of mean link polarization during relaxation processes. The two layers, initially in a paradise state and a two-clique state with same-size cliques, synchronize to one of the three possible matching order states: paradise, a two-clique state with same-size cliques, or a two-clique state with cliques of different sizes (see also Fig. 10).

B. Coupling between different two-clique states

When considering the state where two layers are each in a two-clique state, the overlap between layers may vary between 1 (matching order) and 1/2 (mismatched order). If the cliques match, then the behavior is very similar to the paradise-paradise initial condition, except the mean link polarization is not 1, but depends on the relative sizes of the cliques, down to approximately 0 for cliques of the same size. The critical temperature of transition to disorder is also the same T_c as for the matching paradise state.

If the two-clique states in layers are mismatched, with overlap close to 1/2, then the behavior closely resembles that of the mismatched order of paradise and the two-clique situation

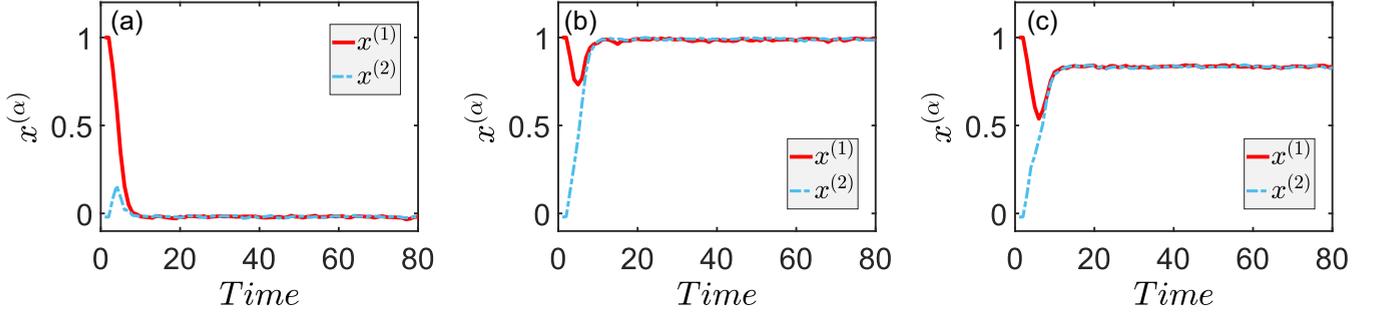


FIG. 10. The final state of the synchronization between two mismatched layers is partially random due to thermal fluctuations. Example trajectories are shown for the synchronization scenario between paradise (red solid line) and two-clique (blue dashed line) layers for $N = 50$, $K = 25$, and $T = 20$. The synchronized state can be (a) a two-clique state similar to the initial condition of the second layer, (b) a paradise state, or (c) a two-clique state of different sizes, matching neither of the two initial layer states. Each shown scenario is a result of independent agent-based simulation under the same initial conditions and parameters.

described in the preceding section. Just as before, there are two critical temperatures, that is, synchronization at T_s and the appearance of disorder at T_d , as well as the same characteristic temperatures T_o above which the system may end up disordered and T_m above which it always falls into disorder instead of synchronizing, as shown in Fig. 11. The exact values of the temperatures T_s , T_o , and T_m may differ, since they are specific to a state, while T_d and T_c are specifically related to disordered and matching order states and thus are fixed and depend only on model parameters.

If the two-clique states are partially matching, then the behavior is qualitatively the same, but values of T_s , T_o , and T_m depend on the overlap o between layers. In the limit of overlap $o \rightarrow 1$, temperature $T_s \rightarrow 0$ and $T_o, T_m \rightarrow T_c$.

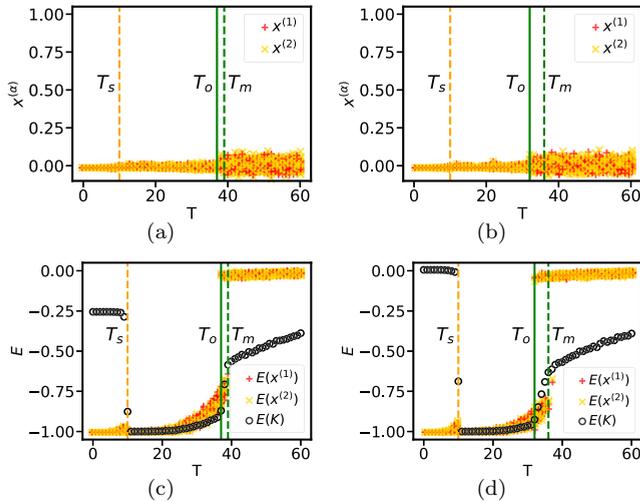


FIG. 11. The behavior of the system of two partially matching [(a) and (c)] and mismatched [(b) and (d)] layers in two-clique states for $N = 50$ and $K = 25$ closely resembles the mismatched order of paradise and two-clique layers seen in Fig. 8. (a) and (b) Variation of mean link polarization for two orthogonal two-clique layers over temperature for 50 independent simulations. (c) and (d) Corresponding variation of mean intralayer (red pluses and yellow crosses) and interlayer (black circles) energies over temperature.

C. Triplex network with mismatched layer pairs

In the case in which the network contains three or more layers, the possible states of the system increase in number. Here we consider a triplex network with one layer in the paradise state and the remaining two layers in two-clique states, which are either matching with $o = 0$ or mismatched with mutual overlap $o = 13/25$ (the same as in the preceding section). The first situation creates two matching layers that are mismatched with the third. The second situation corresponds to every layer being mismatched with the other two. The behavior of the system is similar to the case of a duplex network but features a few important differences.

The most prominent difference is that the two considered situations show different synchronization temperatures $T_s^{(PCC)}$ and $T_s^{(PCC')}$; each state has its own specific synchronization temperature, as can be seen in Fig. 12, showing the simulation results. This is not surprising, as in the case of a state with two matching layers these layers are more stable and exert a stronger, coherent influence on the third, mismatched layer, resulting in a lower temperature $T_s^{(PCC)}$ when the synchronization happens. In the case when all layers are mutually mismatched, there is no coherent impact on any single layer, so their own individual configurations can persist up to a higher transition temperature $T_s^{(PCC')}$. The second detail is how the character of the transition differs between these two situations. The mutually mismatched layer state features symmetry, which means that the transition is similar to the two-layer mismatched state case. None of the layers are stronger, leading to states with varying sizes of the cliques, which may not match any of the initial layers, although it happens less often than in a duplex network. This means that as the number of layers increases, the number of different possible states, and thus state-specific temperatures T_s , T_o , and T_m , increases as well.

D. Phase diagram

Evidence shown in previous sections as well as additional results obtained for a range of parameter K allow us to draw an experimental phase diagram for the system as well as determine possible system states, when they are stable and when they are not, as well as what transitions can occur

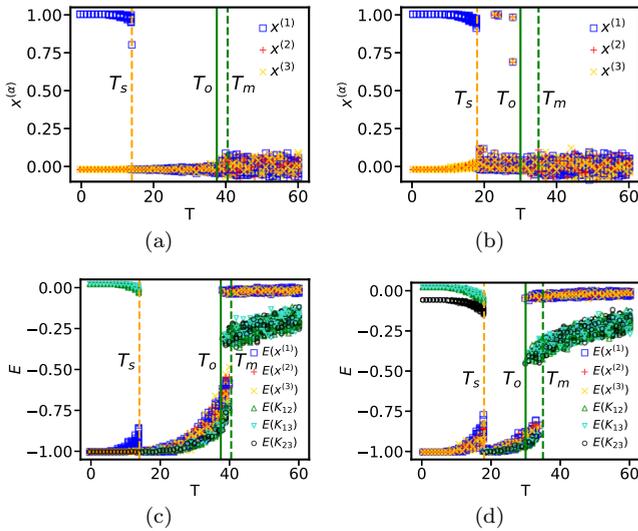


FIG. 12. In a triplex network, there are two classes of states: (a) and (c) two matching layers and a third mismatched (*PCC*) and (b) and (d) all three mutually mismatched layers (*PCC'*). These two state classes show different synchronization temperatures T_s . (a) and (b) Mean polarization (blue squares, red pluses, and yellow crosses) of each of the three layers. (c) and (d) Intralayer energies (blue squares, red pluses, and yellow crosses) as well as interlayer energy (green up triangles, teal down triangles, and black circles) tied to each pair of layers for $N = 50$ and $K = 10$. The *PCC* state consists of the paradise state in one layer and matching two-clique states in the other two; the *PCC'* state consists of the paradise state in one layer and mismatched two-clique states in the other two.

between them. Figure 13 shows the phase diagram for varying temperature and coupling strength between layers. The diagram contains five main parameter regions. In the blue area below T_s any system can exist in either a matching or a mismatched order state. In the green area between T_s and T_d the system may exist only in a matching order state. In the yellow area between T_d and T_c the system may exist in either a matching order or a disordered state. In the red region above T_c the system may exist only in a disordered state. Finally, in the white region between T_d and T_s all of the states, that is, mismatched order, matching order, and disorder, are possible. It is very important to note that while the boundaries set by T_d and T_c , which correspond to the lower bound of disorder and upper bound of matching order, are fixed, the boundary set by T_o is not. Depending on the exact overlap o between mismatched layers, the temperature may vary. If considering a partially matching order state, the qualitative behavior is the same, with temperature T_s varying with overlap o , down to $T_s = 0$ for actually matching layers. The two additional characteristic temperatures T_o and T_m are temperatures above which a system starting in a specific state may become disordered and above which a system always becomes disordered. Similar to T_s , these two characteristic temperatures are also related to a specific initial state and thus depend on the overlap o between layers in the initial state. In Fig. 13 the boundaries shown are for an initial mismatched order state consisting of one layer in the paradise state and one layer in a two-clique state. The characteristic temperatures T_o and T_m increase with overlap, both reaching T_c at overlap $o = 1$ where the layers

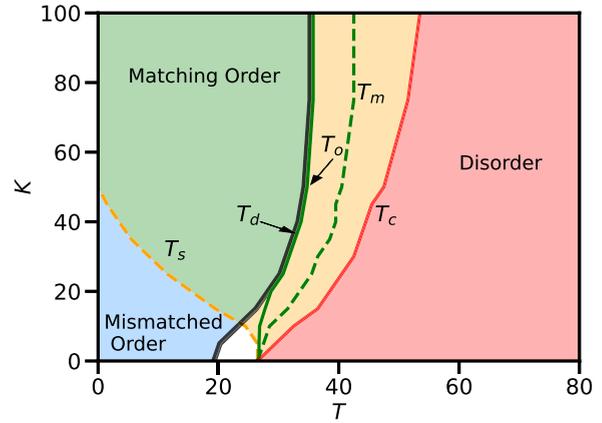


FIG. 13. Phase diagram of the duplex system featuring multistability. The five colored areas are blue (below T_s), a region where both mismatched and matching order states are stable; green (between T_s and T_d), a region where only the matching order state is stable; yellow (between T_d and T_c), a region where both matching order and disordered states are stable; red (above T_c), a region where only the disordered state is stable; and white (between T_d and T_s), a region where mismatched order, matching order, and disordered states are all stable. The boundaries T_d and T_c are fixed, but T_s , T_o , and T_m vary for different states. The results were obtained from agent-based simulations for $N = 50$ and $A = B = 1$ with 50 realizations.

are in fact matching. The temperature T_o here, depicted for a mismatched order state, overlaps with T_d if the coupling strength K between layers is strong enough, meaning that the system may become disordered as soon as the disordered state becomes a stable possibility, but if the coupling strength is lower, then mismatched order can only become disordered at higher temperature $T_o > T_d$.

While the phase diagram in Fig. 13 shows what states are stable under what parameter values, it does not tell us anything about the transitions between them. To understand possible transitions and to better describe characteristic temperatures T_o and T_m , let us look at a fixed K and show the stable states depending on the temperature, as depicted in Fig. 14. Figure 14 shows schematically what states are stable at what temperatures. At low temperatures, both mismatched and matching order states are stable, with intralayer ordering including either a two-clique state or paradise. The partially matching states behave just like mismatched ones and are included in the category but feature transitions at temperatures that depend on the state, most importantly on the overlap o between layers. Temperature T_s is the upper bound for the stability of mismatched order, temperature T_d is the lower bound for the stability of disorder, and temperature T_c is the upper bound for the stability of the matching order state. If the temperature of an already existing matching order state increases above T_c , the system transits into disorder, as depicted with a solid gray arrow in the figure. Similarly, if the temperature for an existing disordered state is lowered below T_d , it transits into a matching order state (also depicted as a solid gray arrow). It is worth noting that the probability that a self-organized matching order state will be a paradise is vanishingly small for larger systems, so the transition in practice is always to a state of matching layers in two-clique

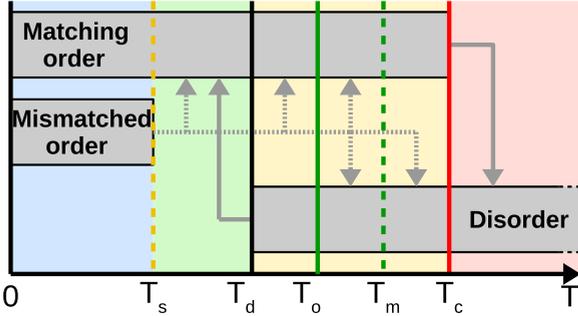


FIG. 14. The Heider balance dynamics on a duplex network is multistable, with different states possible depending on the temperature. Gray boxes show a range of temperatures where a given state is stable for a fixed K value, with gray arrows showing possible transitions between these states. Five characteristic temperatures can be distinguished: synchronization at T_s , the appearance of disorder at T_d , the lower limit of transitions to disorder T_o , the upper limit of transitions to matching order T_m , and the disappearance of order at T_c . Note that the figure is schematic and does not depict actual values for better clarity. The temperature values T_s depend on the exact states they correspond to, since partially matching states with lower overlap have lower temperatures T_s . Characteristic temperatures T_o and T_m depend on the initial state and its overlap between the layers, since higher overlap means higher temperature values, up to T_c for $o = 1$.

states. If the temperature of a system in a mismatched order state is raised above T_s , the transition that happens depends on the new temperature. For $T \in (T_s, T_o)$, it will always become a matching order. For $T \in (T_o, T_m)$, it may become either a matching order or disorder. Finally, for $T > T_m$, the result will always be a disorder, despite matching order being a stable state up until $T_c > T_m$. This means that for $T \in (T_m, T_c)$, while it is possible for an already existing matching order to persist, thermal fluctuations are too large for such a state to appear from a mismatched order. Note that this summary only covers a duplex network with $A^{(1)} = A^{(2)}$. If Heider interaction strengths of layers can be different, the situation may become much more complex.

VI. CONCLUSION

In this paper we explored phase transitions of a multistable thermalized agent system on a multiplex network, with Heider balance intralayer dynamics and Ising interlayer coupling. The system displays several interesting transitions (see Fig. 14). The possible system states include (i) the matching order state, where each layer has the same low-energy ordered state, i.e., the two-clique state or paradise; (ii) the mismatched order state, where each layer is internally ordered but the configuration in each layer is different, resulting in higher interlayer interaction energy (an example of such a state would be a paradise state in one layer and a two-clique state in the second); (iii) the partially matching order state, where each layer is internally ordered into a two-clique state but the split into the cliques only partially match across layers; and (iv) the disordered state, dominated by thermal noise.

We identified three critical temperatures (see Fig. 14).

(i) Above critical temperature T_c , no ordered states are stable. This temperature applies to any number of layers and depends on model and system parameters only. We have cal-

culated a mean-field prediction for T_c for two or more layers. This estimation is in good agreement with numerical results for a two-layered system but is only a rough estimation for systems with more layers. For a system of L layers, the temperature T_c saturates with the strength of positive interlayer coupling, up to $(T_c)_L = L(T_c)_{L=1}$, which has been both predicted by the analytical approach and observed in simulation results. While the mean-field approach only considers fluctuations around the paradise state, numerical results indicate that a state of matching two-clique layers also loses stability above the same T_c .

(ii) At critical temperature T_s , mismatched or partially matching order states exist. This temperature corresponds to a specific state of the system and its value depends on how mismatched the two-clique states in both layers are. Mismatched order and partial matching order states are only stable below this temperature $T_s < T_c$. Above T_s thermal fluctuations allow the system to relax from a higher-energy mismatched order state to a matching order state with a lower energy, similar to the freezing by heating phenomenon [34,35]. If the temperature is high enough, thermal noise may force the system into a high-entropy disordered state instead. For systems with more than two layers, the number of possible states increases, such as two matching and one mismatched layer or three mutually mismatched layers for $L = 3$ total layers. Since each state may feature its own T_c value, there may be multiple different critical temperatures without even considering a semicontinuous spectrum of partial matching between different layers.

(iii) Below critical temperature T_d , the system orders spontaneously. A disordered system at $T < T_d$ spontaneously orders, forming a two-clique matching order state. There is a vanishingly small probability for a larger system to order into a mismatched order state (if $T < T_s$) or for the matching order state to be a paradise, which can be considered a special case of a two-clique state with one clique being of size zero.

In addition, there are two characteristic temperatures T_o and T_m , which correspond to points at which an initially mismatched or partially matching order state starts behaving differently. Temperature T_o is a lower bound for transition to a disordered state, and for an initial mismatched order state and high enough interlayer coupling K it has the same value as T_d . For small K , a disordered state may appear only at $T_o > T_d$. Temperature T_m is an upper bound for the transition from a mismatched or partially matching order to a matching order state. Above T_m the transition from mismatched order or partially matching order state will always be towards a disordered state.

Note that all the critical temperatures T_s , T_d , and T_c , as well as the characteristic temperatures T_o and T_m , depend on model and system parameters, in particular on network size N , intralayer interaction strength A , and interlayer coupling K . In addition, as stated before, critical temperature T_s and characteristic temperatures T_o and T_m are determined for a specific mismatched or a partially matching order state and therefore there are many such temperatures for a given system. The temperature T_c has been predicted analytically (Sec. III) and it is in reasonably good agreement with numerical simulations, but other characteristic temperatures have been identified only through numerical simulations. This is mainly due to these temperatures involving the system layers in two-clique states,

which are outside of the scope of our analytical approach. As a result, any characteristic temperatures involving two-clique states cannot be predicted using our approach. The critical temperatures for the stability of disorder (T_d) and for the stability of order (T_c) are analogous to the critical temperatures observed in [33].

ACKNOWLEDGMENTS

This research received funding from the Polish National Science Center under Alphon Grant No. 2019/01/Y/ST2/00058. This work was funded by the

European Union under the Horizon Europe grant OMINO (Grant No. 101086321). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor European Research Executive Agency can be held responsible for them. This work was cofunded by Polish Ministry of Education and Science with the program International Project Co-Funding. J.A.H. and K.S. additionally acknowledge support from a special grant for leaders of European projects coordinated by the Warsaw University of Technology.

APPENDIX: CRITICAL TEMPERATURE CALCULATIONS FOR A DUPLEX NETWORK

The probabilities of the states \bar{x}_{ij} having energy $E(\bar{x}_{ij})$ described in Eq. (6) are, according to the canonical ensemble (2),

$$P(\bar{x}_{ij}) = \frac{\exp\left\{-\frac{AM}{T}[x_{ij}^{(1)}(x^{(1)})^2 + x_{ij}^{(2)}(x^{(2)})^2]\right\} \exp\left(-\frac{K}{T}x_{ij}^{(1)}x_{ij}^{(2)}\right)}{\sum_{\bar{x}_{mn}} \exp\left\{-\frac{AM}{T}[x_{mn}^{(1)}(x^{(1)})^2 + x_{mn}^{(2)}(x^{(2)})^2]\right\} \exp\left(-\frac{K}{T}x_{mn}^{(1)}x_{mn}^{(2)}\right)}, \quad (\text{A1})$$

where $M = N - 2$ is the number of triads the link ij is part of. Then the expected value $\langle \bar{x}_{ij} \rangle$ [Eq. (4)] will be

$$\langle \bar{x}_{ij} \rangle = \sum_{\bar{x}_{ij}} \frac{\exp\left\{-\frac{AM}{T}[x_{ij}^{(1)}(x^{(1)})^2 + x_{ij}^{(2)}(x^{(2)})^2]\right\} \exp\left(-\frac{K}{T}x_{ij}^{(1)}x_{ij}^{(2)}\right)}{\sum_{\bar{x}_{mn}} \exp\left\{-\frac{AM}{T}[x_{mn}^{(1)}(x^{(1)})^2 + x_{mn}^{(2)}(x^{(2)})^2]\right\} \exp\left(-\frac{K}{T}x_{mn}^{(1)}x_{mn}^{(2)}\right)} \bar{x}_{ij}. \quad (\text{A2})$$

Let us note that \bar{x}_{ij} has only four possible values $\bar{x}_{ij} \in \{(-1, -1), (-1, 1), (1, -1), (1, 1)\}$, so both sums in the equation have a limited number of terms. Then Eq. (A2) becomes

$$\langle \bar{x}_{ij} \rangle = \frac{\exp\left(\frac{2K}{T}\right) \left(\exp\left\{\frac{AM}{T}[(x^{(1)})^2 + (x^{(2)})^2]\right\} - \exp\left\{-\frac{AM}{T}[(x^{(1)})^2 + (x^{(2)})^2]\right\} \right) + \left(\exp\left\{\frac{AM}{T}(-1)^\alpha[(x^{(2)})^2 - (x^{(1)})^2]\right\} - \exp\left\{-\frac{AM}{T}(-1)^\alpha[(x^{(2)})^2 - (x^{(1)})^2]\right\} \right)}{\exp\left(\frac{2K}{T}\right) \left(\exp\left\{\frac{AM}{T}[(x^{(1)})^2 + (x^{(2)})^2]\right\} + \exp\left\{-\frac{AM}{T}[(x^{(1)})^2 + (x^{(2)})^2]\right\} \right) + \left(\exp\left\{\frac{AM}{T}(-1)^\alpha[(x^{(2)})^2 - (x^{(1)})^2]\right\} + \exp\left\{-\frac{AM}{T}[(x^{(2)})^2 - (x^{(1)})^2]\right\} \right)}. \quad (\text{A3})$$

Following mean-field methodology, we assume that the $\langle \bar{x}_{ij} \rangle$ calculated above is the same as the mean-field variable \bar{x} through which the mean field is expressed and the above vector equation can be simplified into a set of self-consistent equations for \bar{x} ,

$$x^{(1)} = f_1(x^{(1)}, x^{(2)}), \quad x^{(2)} = f_2(x^{(1)}, x^{(2)}), \quad (\text{A4})$$

where the right-hand sides can be expressed by

$$f_\alpha(x^{(1)}, x^{(2)}) = \frac{e^{2d} \sinh\{a[(x^{(1)})^2 + (x^{(2)})^2]\} + \sinh\{a(-1)^\alpha[(x^{(2)})^2 - (x^{(1)})^2]\}}{e^{2d} \cosh\{a[(x^{(1)})^2 + (x^{(2)})^2]\} + \cosh\{a[(x^{(1)})^2 - (x^{(2)})^2]\}}, \quad (\text{A5})$$

with $a = \frac{AM}{T}$ and $d = \frac{K}{AM}$. Note that the equations for both components differ only in the sign of the second hyperbolic sine function, expressed here as $(-1)^\alpha$.

The set of self-consistent equations above can be considered a map for the purposes of stability analysis, to determine whether a solution (fixed point) is actually stable and therefore a valid stable state of the system it describes. The Jacobian for the mean-field map [expressed also by Eq. (9)] is

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x^{(1)}} & \frac{\partial f_1}{\partial x^{(2)}} \\ \frac{\partial f_2}{\partial x^{(1)}} & \frac{\partial f_2}{\partial x^{(2)}} \end{bmatrix}, \quad (\text{A6})$$

with the derivatives equaling

$$\frac{\partial f_1}{\partial x^{(1)}} = \frac{4ax^{(1)}e^{2d} \{ \cosh(2d) + \cosh[2a(x^{(2)})^2] \}}{D},$$

$$\frac{\partial f_1}{\partial x^{(2)}} = \frac{2ax^{(2)}(e^{4d} - 1)}{D}, \quad \frac{\partial f_2}{\partial x^{(1)}} = \frac{2ax^{(1)}(e^{4d} - 1)}{D},$$

$$\frac{\partial f_2}{\partial x^{(2)}} = \frac{4ax^{(2)}e^{2d} \{ \cosh(2d) + \cosh[2a(x^{(1)})^2] \}}{D}, \quad (\text{A7})$$

where

$$D = \left(e^{2d} \cosh\{a[(x^{(1)})^2 + (x^{(2)})^2]\} + \cosh\{a[(x^{(1)})^2 - (x^{(2)})^2]\} \right)^2. \quad (\text{A8})$$

Since the two-dimensional maps are symmetric, to simplify calculations we consider $x^{(1)} = x^{(2)} \equiv x$. The fixed point of the mean-field map then becomes

$$x = \frac{\sinh(2ax^2)}{\cosh(2ax^2) + e^{-2d}}, \quad (\text{A9})$$

which in the limit of $d \rightarrow 0$ (no interactions) reduces to a single-layered network, as Eq. (9 a) in [33], and for $d \rightarrow +\infty$ also reduces to the same equation but with the interaction constant $2a$ instead of a . The eigenvalues of the Jacobian

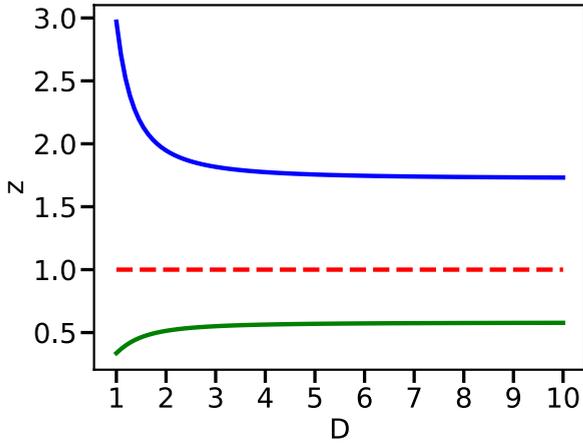


FIG. 15. Equation (A11) has three solutions for the auxiliary variable z that depend only on relative interlayer coupling strength D . Out of them, only the solution $z > 1$ (blue top line) is physically meaningful and corresponds to a positive critical temperature T_c .

matrix J given by Eq. (A7) are

$$\lambda_{+,-} = \frac{4ax[\cosh(2ax^2) + e^{\pm 2d}]}{\{e^d[\cosh(2ax^2) + e^{-2d}]\}^2}, \quad (\text{A10})$$

where λ_+ corresponds to the larger of the two eigenvalues. The critical temperature T_c/AM and x^c can be obtained from a pair of transcendental algebraic relations that describe the fixed point (A9) and criticality condition for the eigenvalue $\lambda_+ = 1$ [see Eq. (A10)]. Multiplying Eq. (A9) by the inverse of Eq. (A10) with $\lambda_+ = 1$ allows us to combine the left-hand x from (A9) and the $4ax$ term from (A10) into a single term $4ax^2$, which means that unknown x only appears as ax^2 . Defining a new variable $z \equiv e^{ax^2}$ and as well as $D \equiv e^d$, we

can write a single equation for z ,

$$8 \ln z = \frac{(z^4 - 1)(z^4 D^2 + 2z^2 + D^2)}{z^2(z^4 + 2z^2 D^2 + 1)}. \quad (\text{A11})$$

This equation has three solutions in general, but only a single relevant solution with $z > 1$. The solution, dependent on D only, is shown in Fig. 15. One can numerically calculate the largest solution z for a fixed value of d . Putting $ax^2 = \ln z$ into Eq. (A9) gives us the link polarization x_c at the critical point

$$x_c = \frac{D^2(z^4 - 1)}{D^2(z^4 + 1) + 2z^2} \quad (\text{A12})$$

and having that, rearranging the definition of z gives us the inverse of a at the critical point

$$\frac{T_c}{AM} = \frac{x^2}{\ln z}. \quad (\text{A13})$$

To understand how the saturation critical temperature relates to a number of layers, we need to define Eq. (A13) at $d \equiv D \rightarrow +\infty$ and at $d = 0$,

$$\frac{T_c}{AM} \Big|_{d \rightarrow +\infty} = \frac{1}{\ln(z_\infty)} \left(\frac{z_\infty^4 - 1}{z_\infty^4 + 1} \right)^2, \quad (\text{A14})$$

$$\frac{T_c}{AM} \Big|_{d=0} = \frac{1}{\ln(z_0)} \left(\frac{z_0^4 - 1}{z_0^4 + 1 + 2z_0^2} \right)^2. \quad (\text{A15})$$

Substituting the values of $z_\infty = 2.97$ and $z_0 = 8.823$ obtained numerically from (A11), the saturation critical temperature is

$$\frac{T_c|_{d \rightarrow +\infty}}{T_c|_{d=0}} = \frac{1.16516 \dots}{0.58258 \dots} = 2 \quad (\text{A16})$$

or

$$T_c|_{d \rightarrow +\infty} \approx 2T_c|_{d=0}. \quad (\text{A17})$$

-
- [1] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, *J. Complex Netw.* **2**, 203 (2014).
- [2] Y. Murase, J. Török, H.-H. Jo, K. Kaski, and J. Kertész, *Phys. Rev. E* **90**, 052810 (2014).
- [3] A. Solé-Ribalta, S. Gómez, and A. Arenas, *Phys. Rev. Lett.* **116**, 108701 (2016).
- [4] M. Szell, R. Lambiotte, and S. Thurner, *Proc. Natl. Acad. Sci. USA* **107**, 13636 (2010).
- [5] S. Boccaletti, G. Bianconi, R. Criado, C. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña Nadal, Z. Wang, and M. Zanin, *Phys. Rep.* **544**, 1 (2014).
- [6] A. I. E. Hosni, K. Li, and S. Ahmad, *Inf. Sci.* **512**, 1458 (2020).
- [7] R. Singh, J. Xu, and B. Berger, *Proc. Natl. Acad. Sci. USA* **105**, 12763 (2008).
- [8] J. Gómez-Gardeñes, I. Reinares, A. Arenas, and L. M. Floria, *Sci. Rep.* **2**, 620 (2012).
- [9] M. Salehi, R. Sharma, M. Marzolla, M. Magnani, P. Siyari, and D. Montesi, *IEEE Trans. Netw. Sci. Eng.* **2**, 65 (2015).
- [10] N. Arinik, R. Figueiredo, and V. Labatut, *Soc. Networks* **60**, 83 (2020).
- [11] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, *Nature (London)* **464**, 1025 (2010).
- [12] C. Atkisson, P. J. Górski, M. O. Jackson, J. A. Holyst, and R. M. D'Souza, *Evol. Anthropol.* **29**, 102 (2020).
- [13] J. Leskovec, D. Huttenlocher, and J. Kleinberg, *Proc. SIGCHI Conf. Human Factors Comput. Syst.* **2**, 1361 (2010).
- [14] F. Heider, *J. Psychol.* **21**, 107 (1946).
- [15] T. Antal, P. Krapivsky, and S. Redner, *Physica D* **224**, 130 (2006).
- [16] T. Antal, P. L. Krapivsky, and S. Redner, *Phys. Rev. E* **72**, 036121 (2005).
- [17] A. Srinivasan, *Proc. Natl. Acad. Sci. USA* **108**, 1751 (2011).
- [18] D. Cartwright and F. Harary, *Psychol. Rev.* **63**, 277 (1956).
- [19] S. A. Marvel, J. Kleinberg, R. D. Kleinberg, and S. H. Strogatz, *Proc. Natl. Acad. Sci. USA* **108**, 1771 (2011).
- [20] F. Rabbani, A. H. Shirazi, and G. R. Jafari, *Phys. Rev. E* **99**, 062302 (2019).
- [21] K. Malarz and K. Kułakowski, *Phys. Rev. E* **103**, 066301 (2021).
- [22] R. Shojaei, P. Manshour, and A. Montakhab, *Phys. Rev. E* **100**, 022303 (2019).
- [23] T. H. Summers and I. Shames, *EPL* **103**, 18001 (2013).
- [24] W. Xia, M. Cao, and K. H. Johansson, *IEEE Trans. Control Netw. Syst.* **3**, 46 (2016).

- [25] M. Yang, X. Wang, L. Ma, Q. He, and M. Huang, *Neural Comput. Appl.* **34**, 16683 (2022).
- [26] G. Facchetti, G. Iacono, and C. Altafini, *Proc. Natl. Acad. Sci. USA* **108**, 20953 (2011).
- [27] P. Doreian and A. Mrvar, *J. Soc. Struct.* **16**, 1 (2015).
- [28] P. J. Górski, K. Kułakowski, P. Gawroński, and J. A. Hołyst, *Sci. Rep.* **7**, 16047 (2017).
- [29] R. P. Kundu and S. Pandey, Cooperation and balance in signed networks: A model of multiplex formation, available at <https://ssrn.com/abstract=4499176> (unpublished).
- [30] E. Cozzo, M. Kivelä, M. D. Domenico, A. Solé-Ribalta, A. Arenas, S. Gómez, M. A. Porter, and Y. Moreno, *New J. Phys.* **17**, 073029 (2015).
- [31] K. Burghardt and Z. Maoz, Imbalance in multiplex networks, available at <https://ssrn.com/abstract=3192890> (unpublished).
- [32] K. Kułakowski, P. Gawroński, and P. Gronek, *Int. J. Mod. Phys. C* **16**, 707 (2005).
- [33] K. Malarz and J. A. Hołyst, *Phys. Rev. E* **106**, 064139 (2022).
- [34] D. Helbing, I. J. Farkas, and T. Vicsek, *Phys. Rev. Lett.* **84**, 1240 (2000).
- [35] J. A. Hołyst, *J. Comput. Sci.* **73**, 102137 (2023).