The Coevolution of Appraisal and Influence Networks leads to Structural Balance

Peng Jia, Noah E. Friedkin, Francesco Bullo

Abstract—In sociology, an appraisal structure, represented by a signed matrix or a signed network, describes an evaluative cognitive configuration among individuals. In this article we argue that interpersonal influences originate from positive interpersonal appraisals and, in turn, adjust individuals' appraisals of others. This mechanism amounts to a coevolution process of interpersonal appraisals and influences. We provide a mathematical formulation of the coevolutionary dynamics, characterize the invariant appraisal structures, and establish the convergence properties for all possible initial appraisals. Moreover, we characterize the implications of our model to the study of signed social networks. Specifically, our model predicts the convergence of the interpersonal appraisal network to a structure composed of multiple factions with multiple followers. A faction is a group of individuals with positive-complete interpersonal appraisals among them. We discuss how this factions-with-followers is a balanced structure with respect to an appropriate generalized model of balance theory.

Index Terms—appraisal evolution, macro-structural models, structural balance theory, mathematical sociology

1 Introduction

Structural balance, a social psychological theory about the network structure of interpersonal appraisals, has attracted attention recently [1], [2], [3] in the studies of political party networks, large-scale online social networks, and cooperation evolution in social networks. Interpersonal appraisal is a ubiquitous natural relation of evaluative (positive or negative) cognitive orientation of one individual toward another. Cartwright and Harary's seminal work [4] on the signed digraph formalization of Heider's analysis [5], [6] of balanced cognitive configurations is now referred to as the *classical model of structural balance*. This model posits the existence of tensions corresponding to configurations of appraisals among three individuals.

Based on empirical observations, a *generalized model of structural balance* is introduced by Davis and Leinhardt [7] and studied by Johnsen [8]: the tension assumptions are relaxed and more complex realizations of structural balance are allowed. While structural balance theory typically focuses on the static appraisal networks, recent research [9], [1], [3] has concentrated on dynamical models of structural evolution. In what follows, we first review some relevant literature and later state our problem of interest.

Static Structural Balance Theory: In structural balance theory, a complete signed matrix $X \in \{-1,+1\}^{N\times N}$, which we call the *appraisal matrix*, represents the interpersonal ties in a group of N individuals, where x_{ij}, $i,j \in \{1,\dots,N\}$, equals to $+1$ if individual i has a positive appraisal of individual j, and x_{ij} equals to -1 if i has a negative appraisal of individual j. The matrix X is used to describe the interpersonal appraisal structure of the group. Any dyad $\{i,j\}$ in the group associated with X has three possible types: mutual (M), asymmetric (A), or null (N). A dyad $\{i,j\}$ is M-related if $x_{ij} = x_{ji} = 1$; it is N-related if $x_{ij} = x_{ji} = -1$; and it is A-related otherwise (i.e., $x_{ij} \times x_{ji} = -1$). Consequently, there are 16 different types of *triads* for an appraisal structure.

The classical model of structural balance posits that an appraisal structure is balanced if the following four statements by Heider [6] are satisfied: "my friend's friend is my friend," "my friend's enemy is my enemy," "my enemy's enemy is my friend," and "my enemy's friend is my enemy." Mathematically, a signed digraph is balanced under these conditions if the value of all cycles (i.e., closed paths beginning and ending with the same node) are positive with respect to the product of all edges of the cycle. A remarkable implication of the classical theory of structural balance is that an appraisal network is tension-free (balanced) if and only if it is positive-complete or partitioned into two positive-complete subgraphs. A positive-complete subgraph, also referred to as a *faction*, is a social subgroup in which each individual has a positive appraisal of each individual in the faction and a non-positive appraisal of any other individual.

The generalized model of structural balance is proposed by Davis and Leinhardt [7] and Johnsen [10], [8]. In this theory, a *micro-model* specifies which subset of the 16 triad types is permitted and, therefore, which resulting structural networks (signed digraphs containing only permitted triad types) are admissible. The

This work is supported by the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office and by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF-15-1-0577. The content of the information does not necessarily reflect the position or the policy of the Government; no official endorsement should be inferred.
Peng Jia, Noah Friedkin, and Francesco Bullo are affiliated with the Center for Control, Dynamical Systems and Computation, UCSB. Email: {pjia,bullo}@engineering.ucsb.edu, friedkin@soc.ucsb.edu

resulting structural networks are referred to as *macro-structural models*. In contrast with the classical model, the generalized model of structural balance (in many of its micro-model realizations) allows for arbitrary numbers of factions and directed acyclic graphical structures among them. We review this theory in Section 2 and refer to [8] for a detailed treatment.

Dynamical Models for Social Balance: Classical structural balance theory and its generalizations do not specify the mechanisms that transform unbalanced appraisal networks into balanced networks. However, as Marvel et al. [1] noted "... 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."

In Kułakowski's work [9] a continuous-time model of structural balance was presented: given an appraisal matrix $X \in \mathbb{R}^{N\times N}$, the dynamical system

$$dX/dt = X^2 \tag{1}$$

governs the evolution of the appraisal structure over time. Here, the entry x_{ij} denotes the appraisal of individual j held by individual i. Numerical simulations showed that, for almost all initial $X(0)$, the system reaches the structural balance postulated by the classical model in finite time. Kułakowski proved the convergence to this structural balance for an $N = 3$ network. Marvel et al. [1] proved that for a random initial symmetric matrix $X(0)$, the classical structural balance is obtained by the dynamical system (1) in finite time with a probability converging to 1 as the population size goes to infinity. Traag et al. [3] extended the convergence results to normal initial matrices (X is normal if $X^TX = XX^T$). However, for generic initial appraisal matrices, the convergence to structural balance is not necessarily observed [3] in the dynamical system (1). Another interesting continuous-time dynamical model presented by Traag et al. is

$$dX/dt = XX^T. \tag{2}$$

For this model, there exists [3] a dense set of initial conditions $X(0) \in \mathbb{R}^{N\times N}$, such that a balanced structure is achieved generically in finite time. In the first model (1), the appraisal of the individual j held by the individual i is adjusted based on the appraisals of j held by all other individuals. In the second model (2), the appraisal of the individual j held by the individual i is adjusted based on the appraisal of all individuals held by j. In other words, the second model features a cooperative behavior in the appraisal evolution: each individual tends to befriend other individuals who think alike.

The literature on social structural balance also includes a research stream on social energy landscape [4], [11], [12], [2]. These works are motivated by studies which suggest that certain triad types are more stable than others. In these works, an energy landscape is defined to describe structural balance. Numerical experiments show that energy landscapes often feature local minima called jammed states [11], [12]. Antal et al. [11] consider a discrete-time dynamical model, where the signs of the edges of an appraisal network are optimally adjusted under the constraint of a monotonic increase of balanced triads. It is shown that the assumption of an optimal monotonic increase of balanced triads does not suffice to generate the structural balance of the classical model.

In summary, we argue that it is of considerable interest to postulate social psychological mechanisms of appraisal dynamics and to characterize the conditions under which they present outcomes consistent with structural balance theory.

The Coevolution of Interpersonal Appraisals and Influences: We propose a novel sociological model for the evolution of interpersonal appraisals towards generalized structural balance. Our approach treats the evolution of interpersonal appraisals as a special case of opinion dynamics: the evolving opinions are the individuals' bundles of signed cognitive appraisals towards the other individuals. In other words, we ground dynamic structural balance theory in the theory of opinion dynamics and influence networks. As opinion and appraisal evolution mechanism, we adopt the widely-established DeGroot averaging model [13]. Notably, our dynamical model predicts general numbers of factions and rich structure among faction and is consistent with a particular micro-model from generalized balance theory. Our work is also motivated by the coevolution process [14] in which an influence network is associated with an appraisal network, interpersonal influences adjust individuals' appraisals, and these adjusted appraisals lead to an adjusted influence network.

With this background motivation, we consider a discrete-time dynamical model of structural balance, where the dynamics combine both appraisal and influence structure evolution. For a group of $N \geq 2$ individuals with initial appraisals $X(0) \in \mathbb{R}^{N\times N}$, the evolution of the appraisal matrix $X(t)$ is determined by a discrete-time DeGroot averaging model:

$$X(t+1) = W(X(t))X(t), \quad t = 0, 1, 2 \ldots. \tag{3}$$

Here, the row-stochastic *influence matrix* $W(X(t))$ depends on the state $X(t)$ as follows: interpersonal influences for each individual are *proportional* to the *positive* appraisals accorded to the individual by all other individuals. In other words, the entry $w_{ij}(X)$, $i \neq j$, is proportional to x_{ij} if x_{ij} is positive and zero otherwise. We provide a detailed mathematical definition in the modeling section below. As the dynamical processes of interpersonal appraisals and influences are interdependent, we refer to (3) as the *coevolution model of interpersonal appraisal and influence*.

Our setup features several differences with the existing dynamical models on structural balance (e.g., the models (1) and (2)). First, our model (3) considers both appraisal and influence evolution and a novel process to generate and adjust the interpersonal influence network. Note that the appraisal structure plays the role of the influence structure in the previous work [9], [1], [3]. Second, in our model appraisals are modified over time as convex combination of existing appraisals; this guarantees that the appraisals never diverge (by comparison, divergence occurs in the models of (1) and (2)). Third, while positive and negative appraisals with heterogeneous strengths are allowed in our model, our basic assumption is that only positive interpersonal appraisals translate into interpersonal influences. Accordingly, we expect the evolutions of the proposed coevolution model to asymptotically satisfy two statements in the classic balance model: "my friend's enemy is my enemy" and "my friend's friend is my friend." (On the contrary, there are a priori no reasons why the other two structural balance theory statements, "my enemy's enemy is my friend" and "my enemy's friend is my enemy," should be satisfied by the evolutions of our model.) By comparison, the models (1) and (2) satisfy all four statements in the classic balance model and predict only one or two factions of structural balance given certain initial conditions. In other words, the classical model and the models in equations (1) and (2) are therefore not predictive of the situations, empirically observed in [7], in which more than two factions are often observed. Moreover, for generic initial conditions, the solution of the model (1) converges to a rank-1 matrix which is, in general, not structurally balanced, even in the language of the generalized structural balance. Finally, it is noted that our model (3) is a discrete-time dynamical model, while the equations (1) and (2) are in continuous time.

Contributions: We propose and analyze the novel coevolution model of interpersonal appraisal and influence given in equation (3). As a first step, we provide an explicit and concise mathematical formulation for this discrete-time nonlinear system. As the main set of contributions, we study the asymptotic convergence and equilibrium properties of this nonlinear coevolution system. We provide a mathematical analysis on the structural position properties of the individuals; we predict the equilibrium appraisals for individuals in a sink, intermediate or source strongly connected component (SCC). We claim that (1) all individuals in a sink SCC of the equilibrium positive digraph will reach an appraisal consensus on each individual, (2) all intermediate SCCs vanish in the dynamical equilibrium, (3) each source SCC is a singleton and the appraisal of individuals in source SCCs are determined by the appraisals held by the individuals from the sinks, and (4) all equilibrium appraisal networks have a factions-followers-outsiders structure. Here, a faction is a sink SCC with positive appraisals within the component; a follower is a singleton source SCC; and an outsider is an isolated singleton sink SCC with a non-positive self-appraisal.

As the second set of contributions, we demonstrate that all invariant macro-structures are the special cases of the factions-followers-outsiders structure. This finding is related to the concept of core-periphery structure, a prevalent notion in world systems [15], economics [16], and social networks [17], [18]. In other words, the factions-followers-outsiders structure exhibits the properties of a multi-core-periphery structure: dense, cohesive cores (factions) and sparse, unconnected peripheries (followers or outsiders). Furthermore, all equilibrium appraisal structures via the coevolution are discussed and the exclusive set of macro-structural models are predicted to be structurally balanced under our model.

Lastly, we illustrate our results by numerical simulations. In particular, we inspect the different convergence and equilibria performances for different self-influence parameters.

These findings are of sociological interest in their advancement of the dynamical foundations of structural balance in social groups. Our rigorous results for the coevolution model of interpersonal appraisal and influence suggest that interpersonal appraisal networks evolve toward a set of structural equivalent bundles and predict the stable macro-structures under the coevolution model. These findings contribute to the rapidly-growing research on coevolutionary networks [19], [20], to the literature on social network formation and coordination games [21], [22] and, more broadly, to the study of complex networks and evolutionary rules [23], [24].

Finally, we note that the interesting attractor topologies admitted by the generalized balance theory in this paper have important practical implications on the capacity of interpersonal influence systems to resolve social conflicts and disseminate innovations. The hierarchical topological attractors we characterize provide a basis of population consensus generation and diffusion of innovations. The theory suggests that a small set of influential factions, and the chains of positive appraisals that link other followers to them, determine the beliefs, opinions, and behaviors of large numbers of individuals on a variety of issues.

Organization: The rest of the paper is organized as follows. Section 2 introduces some notation and preliminary concepts. Section 3 describes the coevolution model. Sections 4, 5 and 6 discuss, respectively, the topological, asymptotic and structural balance properties of our model. Section 7 contains our conclusions. All technical proofs are in the Appendices.

2 Preliminary Concepts

For a vector $x \in \mathbb{R}^n$, we let $x \geq 0$ and $x > 0$ denote component-wise inequalities. We adopt the abbreviations $\mathbb{1}_n = [1, \ldots, 1]^T$ and $\mathbb{0}_n = [0, \ldots, 0]^T$. Given a column vector $[x_1, \ldots, x_n]^T \in \mathbb{R}^n$, we let $\mathrm{diag}(x)$ denote the diagonal $n \times n$ matrix whose diagonal entries are $x_1, \ldots, x_n$. For signed matrix pattern analysis, we let "$-$" represent a block matrix with *non-positive* entries and let "$+$" represent a block matrix with positive entries for the simplicity of presentation. If two matrices $A, B \in \mathbb{R}^{N \times M}$ have the same positive/non-positive entry pattern, we denote $A \sim B$. For example, one matrix A with all positive entries or all non-positive entries is denoted by $A \sim [+]$ or $A \sim [-]$, respectively. For $x \in \mathbb{R}$ and $A = [a_{ij}] \in \mathbb{R}^{N \times M}$, we write $x^+ = \max\{x, 0\} \in \mathbb{R}_{\geq 0}$ and $A^+ := [\max\{a_{ij}, 0\}] \in \mathbb{R}^{N \times M}_{\geq 0}$.

Elements of graph theory: An *undirected graph* (in short, *graph*) is an ordered pair $G = (V, E)$, where V is a set of nodes and E is a set of unordered pairs of nodes. A *directed graph* (in short, *digraph*) $G = (V, E)$ consists of a node set V and a set E of ordered pairs of nodes, i.e., $E \subset V \times V$. For $i, j \in V$, the ordered pair (i, j) denotes a directed *edge* from i to j, where i is called an *in-neighbor* of j, and j is called an *out-neighbor* of i. The *in-degree* and *out-degree* of j are the numbers of in-neighbors and out-neighbors of j, respectively. Every node of in-degree (resp. out-degree) 0 is called a *source* (resp. *sink*). A node with both non-zero in-degree and out-degree is an *intermediate* node and a node with both 0 in-degree and out-degree is an *isolated* node.

A *directed path* in a digraph G is an ordered sequence of nodes such that any pair of consecutive nodes in the sequence is a directed edge. A directed path is *simple* if no node appears more than once in it, except possibly for the initial and final node. G is *strongly connected* if there exists a directed path from any node to any other nodes. A node of G is *globally reachable* if it can be reached from any other node by traversing a directed path. A *cycle* in G is a simple directed path that starts and ends at the same node. A *directed acyclic graph* (*DAG* in short) is a digraph that has no cycles. G is *aperiodic* if there exists no integer $k > 1$ dividing the length of each cycle in G. Given a digraph $G = (V, E)$, a linear ordering of nodes is an *inverse topological sorting* if, for any edge $(i, j) \in E$, j precedes i in the ordering. Any DAG has one inverse topological sorting, which may not be unique [25].

A digraph (V', E') is a *subgraph* of (V, E) if $V' \subset V$ and $E' \subset E$. A subgraph H is a *strongly connected component* (*SCC* in short) of G if H is strongly connected and any other subgraph of G strictly containing H is not strongly connected. The *condensation digraph* of G, denoted $C(G)$, is a digraph whose nodes are the SCCs of G and in which there exists a directed edge from the SCC H_1 to the SCC H_2 if and only if there exists a directed edge in G from a node of H_1 to a node of H_2. Each $C(G)$ is a DAG and has at least one sink and one source. We say that H_1 *is directly connected to* H_2 in G if there exists a directed edge in $C(G)$ from H_1 to H_2.

Elements of matrix theory: A non-negative matrix is *row-stochastic* if all its row sums are equal to 1. Given a square non-negative matrix $M = \{m_{ij}\}_{i,j \in \{1,\ldots,n\}}$, its *associated digraph* $G(M)$ is the digraph with node set $\{1, \ldots, n\}$ and with edge set defined as follows: (i, j) is a directed edge if and only if $m_{ij} > 0$. M is *irreducible* if $G(M)$ is strongly connected; M is *reducible* if it is not irreducible. M is *aperiodic* if $G(M)$ is aperiodic. M is *primitive* if there exists $k \in \mathbb{N}$ such that M^k is a positive matrix. It is known (e.g., see Example 1.2 in [26]) that M is primitive if and only if M is irreducible and aperiodic.

Generalized models of structural balance: Micro-models and macro-structural models were introduced in Johnsen [8] (see also [27, Section 8.3]) to generalize the structural balance models studied in [4], [6]. A *micro-model* is defined to be a subset of the 16 possible triad types. Associated with a particular micro-model, a *macro-structural model* (or a *macro-structure* in short) is defined to be the set of networks containing only the triad types in the micro-model. We call such a pair of the macro-structure and the micro-model *consistent* and, equivalently, we say that a set of macro-structure networks is *structurally balanced* with respect to a specified micro-model.

We rephrase and generalize the dyad types of interpersonal appraisals (the classical version of which was introduced in [4] and briefly presented in the introduction) as follows: dyad $\{i, j\}$ are *M-related* if $x_{ij} > 0$ and $x_{ji} > 0$; they are *N-related* if $x_{ij} \leq 0$ and $x_{ji} \leq 0$; and they are *A-related* otherwise (i.e., one of $\{x_{ij}, x_{ji}\}$ is strictly positive and the other is non-positive). The triad types are then generalized by the rephrased dyad types such that the values of the appraisal relations are not constrained to -1 or 1. An *M-clique* is a set of individuals who are completely connected by M-links (i.e., links with M-relations).

3 Model of appraisal and influence coevolution

In this section, we present a dynamical model which investigates how an influence network may emerge from interpersonal appraisals in a social network and how the appraisal relations may be modified by interpersonal influences. We adopt the term "structure" for the non-positive/positive (or zero/non-zero, respectively) pattern of the appraisal (or influence, respectively) relations among the network irrespective of their values. We adopt the term "matrix" to describe the exact quantification of the interpersonal appraisals (or influence weights, respectively). In other words, an appraisal structure is a set of appraisal matrices with a certain sign pattern.

We consider a group of $N \geq 2$ individuals with interpersonal appraisals represented by a signed matrix $X \in \mathbb{R}^{N \times N}$. Each entry $x_{ij}, i, j \in \{1, \ldots, N\}$, of the *appraisal matrix* X represents individual i's *interpersonal appraisal* of individual j. Additionally, we allow the entries of X equal to 0: $x_{ij} = 0$ implies that individual i has neither positive nor negative appraisal of individual j, or that i does not know j. Thus, we relax the complete digraph or weakly connected digraph assumption on appraisal structures, which was widely adopted in the previous work (e.g., see [4], [1], [3]).

Individuals' appraisals, i.e., signed evaluative orientations of particular strengths, are often automatically generated without conscious effort [28], [29], and these appraisals are important antecedents of displayed cognitive and behavioral orientations toward objects [30]. The available empirical evidence is also consistent with the assumption that individuals update their appraisals as convex combinations of their own and others' displayed appraisals. This convex combination is based on weights that are automatically generated by individuals in their responses to the displayed appraisals held by other individuals. This specification appeared in the literature on opinion dynamics in the early works by French [31], Harary [32], and DeGroot [13]. Especially, in Anderson's seminal information integration theory [33], the convex combination mechanism was regarded as a fundamental "cognitive algebra" of the mind's synthesis of heterogeneous information. Therefore, in this article, we formulate individuals' appraisals about others by a trajectory $t \mapsto X(t)$ that is determined by the seminal DeGroot averaging model:

$$X(t+1) = W(t)X(t), \quad t \geq 0, \tag{4}$$

with initial appraisals $X(0) \in \mathbb{R}^{N\times N}$, and with a sequence of *influence matrices* $\{W(t)\}_{t\geq 0}$. Here, each influence matrix $W(t)$ is assumed to be non-negative and its entry $w_{ij}(t)$, $i, j \in \{1, \dots, N\}$, represents the *interpersonal influence weights* that the individual i accords to individual j at time t.

Our analysis of appraisal evolution (4) depends only on the influence matrices $\{W(t)\}_{t\geq 0}$. The key feature of the proposed model is that $W(t)$ is determined by the appraisal matrix $X(t)$ at each time t. Motivated by Friedkin and Johnsen's work [14], we associate an influence matrix to an appraisal matrix as follows: (i) the interpersonal influence $w_{ij}(t)$ is strictly positive precisely when the corresponding appraisal $x_{ij}(t)$ is strictly positive, (ii) non-positive appraisals lead to zero interpersonal influence weights, and (iii) each individual has a positive self-weight in the influence network. Specifically, given a small *self-appraisal constant* $\epsilon > 0$, individuals' interpersonal influences are determined as functions of the interpersonal appraisals X by

$$w_{\epsilon,ij}(X) = \begin{cases} \dfrac{1}{\sum_{j=1}^{n} x_{ij}^{+} + \epsilon}(x_{ij}^{+} + \epsilon), & \text{if } j = i, \\ \dfrac{1}{\sum_{j=1}^{n} x_{ij}^{+} + \epsilon} x_{ij}^{+}, & \text{if } j \neq i. \end{cases} \tag{5}$$

An equivalent definition of $W_\epsilon(X) = [w_{\epsilon,ij}]$ is:

$$W_\epsilon(X) = \operatorname{diag}\big((X^{+} + \epsilon I_N)\mathbb{1}_N\big)^{-1}\big(X^{+} + \epsilon I_N\big). \tag{6}$$

By continuity, we also define $W_0(X) = \lim_{\epsilon\to 0^+} W_\epsilon(X)$. Note that, if each entry of the i-th row of X is negative or zero, then $w_{\epsilon,ii}(X) = 1$. This equality implies that the i-th individual assigns zero weight to all other individuals and that his appraisals are therefore unchanged after one step of the evolution.

Definition 3.1 (Appraisal and influence coevolution model)**.** *Given a group of $N \geq 2$ individuals, let $X(t) \in \mathbb{R}^{N\times N}$ be the appraisal matrix and $W(t) \in [0,1]^{N\times N}$ be the influence matrix at time $t \geq 0$. For $t \in \mathbb{Z}_{\geq 0}$, the* interpersonal appraisal and influence coevolution system *is defined by:*

(i) Appraisal evolution:

$$X(t+1) = W(t)X(t),$$

(ii) Influence evolution:

$$W(t) = W_\epsilon(X(t)).$$

For simplicity of presentation, the model assumes the positive parameter ϵ to be homogeneous among the individuals. Nevertheless, this assumption is not necessary and is relaxed in Section 5. Fig. 1 illustrates our proposed coevolution model as in Definition 3.1.

Lemma 3.1 (Properties of influence matrices)**.** *For any $\epsilon \geq 0$ and any appraisal matrix $X \in \mathbb{R}^{N\times N}$, the influence matrix $W_\epsilon(X)$ is row-stochastic and, if $\epsilon > 0$, aperiodic.*

4 Topological Properties of the Coevolutionary Dynamics

In this section we study the topological properties of the coevolution model of interpersonal appraisal and influence (4)-(5), and focus on the long-term connectivity properties of the graphs describing interpersonal influences.

We call the digraph $G(W)$ associated to the influence matrix W the *influence digraph*, and call the digraph $G^{+}(X)$ associated to X the *positive (appraisal) digraph* for which a directed edge (i, j) exists if and only if $x_{ij} > 0$. It is clear that the adjacency matrix X^{+} of $G^{+}(X)$ has the same positive/non-positive entry pattern as $W(X)$ except possibly for the diagonal entries. Consequently, the two digraphs $G^{+}(X)$ and $G(W)$ have the identical set of nodes and the identical set of edges except possibly for self-loops. In what follows we analyze the evolution of $G^{+}(X(t))$ along the trajectory $X(t)$ of the dynamical system (4)-(5).

Since $C(G)$ is a DAG, by relabelling its nodes from an inverse topological sorting, the adjacency matrix of

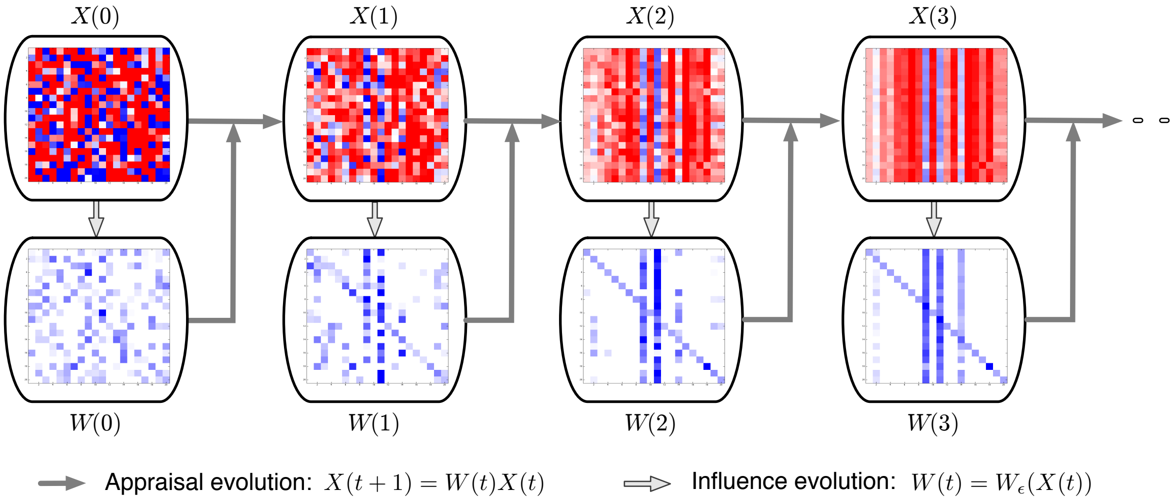


Fig. 1. Coevolution of appraisal and influence: the colormaps correspond to the signed appraisal matrices and influence matrices. In these colormaps and what follows red colors are negative appraisals and blue colors are positive appraisals or influences. The various color depths represent different appraisal or influence values.

$C(G)$ is a block lower triangular non-negative matrix as follows

$$A = \begin{bmatrix} D_1 & 0 & \dots & 0 \\ S_{21} & D_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \dots & D_n \end{bmatrix}, \tag{7}$$

where n nodes $\{H_i\}_{i\in\{1,\dots,n\}}$ exist for the condensation digraph $C(G)$. The matrix A can also be regarded as an adjacency matrix of G if we look at D_i and S_{ij} as block matrices and $\{H_i\}_{i\in\{1,\dots,n\}}$ represent the SCCs of G. The ordering of nodes within an SCC H_i is inessential.

As the topology of positive appraisal digraph or influence digraph may vary along the coevolution, we denote the nodes of $C(G^+(X(t)))$ or the SCCs of $G^+(X(t))$ as $\{H_i(t)\}_{i\in\{1,\dots,n(t)\}}$, where $n(t)$ denotes the numbers of SCCs associated with the positive digraphs. With a slight abuse of notation, we refer to $H_i(t)$ as both a SCC of $G^+(X(t))$ at time t and the subset of nodes of $G^+(X(t))$ belonging to that SCC. That is, $H_i(t)$ may represent the same subset of nodes in the digraph $G^+(X(t+1))$ as those in the digraph $G^+(X(t))$ forming the SCC $H_i(t)$ even though $H_i(t)$ may not be an SCC of $G^+(X(t+1))$ any more.

Theorem 4.1 (Finite-time stability of the SCCs of positive digraphs)**.** *Let $X(t)$ be a trajectory of the coevolution system* (4)-(5) *with $\epsilon \geq 0$. Pick a time $t \in \mathbb{Z}_{\geq 0}$ and perform an inverse topological sorting of the condensation digraph $C(G^+(X(t)))$. For any two nodes $H_i(t)$ and $H_j(t)$ with labels $i < j$ in $C(G^+(X(t)))$,*

(i) no directed edge can appear from a node of $H_i(t)$ to a node of $H_j(t)$ in $G^+(X(t+1))$; and

(ii) if $C(G^+(X(t)))$ contains no directed path from $H_j(t)$ to $H_i(t)$ with length 1 or 2, then no directed edge can appear from a node of $H_j(t)$ to a node of $H_i(t)$ in $G^+(X(t+1))$.

Therefore,

(iii) no two SCCs of $G^+(X(t))$ can merge at time $t+1$ (whereas an SCC of $G^+(X(t))$ may split into multiple SCCs at time $t+1$);

(iv) the number $n(t)$ of SCCs of $G^+(X(t))$ is non-decreasing; and

(v) there exists a finite time τ such that the SCCs of $G^+(X(t))$ remain unchanged for all subsequent times $t \geq \tau$.

As $\{H_i(t)\}$ and $n(t)$ remain unchanged for all $t \geq \tau$, we denote $H_i = H_i(\tau)$ and $n = n(\tau)$ for simplicity in the following discussions. It is noted that the blocks below the diagonal in (7) are varying via the coevolution system (4)-(5) even after time τ, and therefore, the directed edges from the node H_i to the node H_j for all $i > j$ do not necessarily remain unchanged in the condensation digraphs $C(G^+(X(t)))$ for $t \geq \tau$. That is, the topology evolution of the positive digraphs may not stabilize at τ. Moreover, due to the discontinuity of $G^+(X(t))$, although the SCCs of $G^+(X(t))$ are unchanged after some finite time τ, they are not necessarily equal to the SCCs of $G^+(\lim_{t\to\infty} X(t))$.

Fig. 2 illustrates one example showing the appraisal digraph evolution and the edge evolution as described in Theorem 4.1. We may verify the claims of the theorem by this example: Given $N = 3$ and $\epsilon = 0.5$, we observe that (1) the positive appraisal digraph $G^+(X(0))$ has only one SCC but this SCC splits into three SCCs at time $t = 1$, i.e., singleton SCC nodes $H_1(1), H_2(1), H_3(1)$; (2) no directed edge can appear between the two SCCs $H_2(t)$ and $H_3(t)$, and no edge can appear from $H_2(t)$ or $H_3(t)$ to $H_1(t)$ for $t \geq 1$; (3) the three SCCs $H_1(t), H_2(t), H_3(t)$ never merge for all $t \geq 1$ and the number of the SCCs are non-decreasing. By simple calculation, we know that the stability time for both the SCC evolution and topology evolution of the positive digraphs considered in Fig. 2 is $\tau = 1$ and the number of the stable SCCs is $n(\tau) = 3$.

5 Asymptotic properties of the coevolutionary dynamics

In this section we study the asymptotic convergence properties of the coevolution model of interpersonal appraisal and influence (4)-(5). Both analytic and numerical results are presented.

5.1 Theoretical results

We start with arbitrary initial conditions and subject to a non-zero assumption on ϵ and we show that each trajectory converges asymptotically to an equilibrium matrix and we characterize the structure of the equilibrium matrices. Second, we discuss sufficient and necessary conditions such that certain appraisal structures observed in finite time remain unchanged in infinite-time limit.

Definition 5.1 (Factions-followers-outsiders structure)**.** *An appraisal matrix X has a* factions-followers-outsiders structure *if each strongly-connected component of $G^+(X)$ is either*

(i) a sink in $C(G^+(X))$, called a faction*, composed of an arbitrary number of nodes, all of which are completely connected and have positive self-loops in $G^+(X)$; or*

(ii) a source in $C(G^+(X))$, called a follower*, composed of a single node with directed edges pointing to each node in one or more factions and without self-loop in $G^+(X)$; or*

(iii) an isolated node in $C(G^+(X))$, called an outsider*, composed of a single node in $G^+(X)$ with zero in-degree, zero out-degree and no self-loop.*

The node in a follower component with directed edges towards one or more factions is called a *follower of that or those factions*. If there are only factions and followers associated with $G^+(X)$, it is called a *factions-followers structure*. This definition is illustrated in Figure 3: note that a faction can have one or multiple nodes and can have zero, one or more followers. A faction-follower-outsiders structure may includes one or multiple copies of the set or a subset of the structures shown in figure.

Theorem 5.1 (Asymptotic appraisal matrices)**.** *For the coevolution system* (4)-(5) *with $\epsilon > 0$, each trajectory $X(t)$ converges asymptotically to an equilibrium X^* (function of $X(0)$ and ϵ). Moreover, each equilibrium matrix X^* has the following properties:*

(i) X^* has a factions-followers-outsiders structure;

(ii) for a faction of $G^+(X^*)$, the appraisals of each individual in the network held by all individuals in the faction are the same: this faction's appraisal of one individual is positive if the individual belongs to this faction and is non-positive otherwise;

(iii) for a follower of $G^+(X^*)$, the appraisal of each individual in the network held by the follower is a convex combination of the appraisals of that individual held by the factions the follower follows. In particular, if the follower follows only one faction, then its appraisal of each individual in the network is identical to that held by the faction;

(iv) for an outsider of $G^+(X^*)$, the appraisal of each individual in the network held by the outsider is non-positive.

Theorem 5.1 says that, subject to the coevolution (4)-(5), the factions-followers-outsiders structure is the only possible equilibrium structure of the appraisal matrix X^*. In particular, we know that: (i) the rows of X^*, corresponding to all individuals' appraisals in a faction, are identical and are equal to $\mathbb{1}_{N_s} v^T$ for some $v = [v_j] \in \mathbb{R}^N$, where N_s is the cardinality of the faction, $v_j > 0$ if node j is in the faction and $v_j \leq 0$ otherwise; (ii) all followers, i.e., source strongly-connected components, are singletons and their appraisals are determined by the appraisals held by the factions, of which the followers hold positive-complete appraisals in the equilibria; and (iii) all outsiders have non-positive appraisals of each individual in the group. The examples of the convergence in Theorem 5.1 are presented in Fig. 4-6. The nodes of the right graphs are M-cliques. $M_{(1)}$ in these figures is a special M-clique with only one node. More discussions of M-clique are referred to Section 6.

Next, we analyze when and what finite-time structures (i.e., strongly-connected components of the positive digraph) remain unchanged in the asymptotic limit. For simple presentation, we denote the matrix corresponding to interpersonal appraisals in a sink SCC H_s by $X_s \in \mathbb{R}^{N_s \times N_s}$, where N_s is the node cardinality of the SCC.

Theorem 5.2 (Finite-time properties determining asymptotic structures)**.** *For the coevolution system* (4)-(5) *with $\epsilon > 0$, let the trajectory $X(t)$ satisfy $X^* = \lim_{t\to\infty} X(t)$. Then*

(i) a sink SCC H_s of $G^+(X(t))$ exponentially converges to a faction in X^* if and only if there exists a time $t_1 \geq t$ such that $X_s(t_1)$ has one column with all positive entries;

(ii) $G^+(X^*)$ has a globally reachable node if and only if there exists a time t such that $X(t)$ has one column with all positive entries;

(iii) if all entries of $X(t)$ are non-negative and $G^+(X(t))$ is irreducible, then $G^+(X^*)$ is one faction and all individuals have the same appraisal of each individual in the group; and

(iv) an outsider of $G^+(X(t))$ remains an outsider for all following times and in the limit $G^+(X^*)$.

The statement (ii) extends the statement (i) to an appraisal structure of which the positive digraph is at least weakly connected and has only one sink SCC.

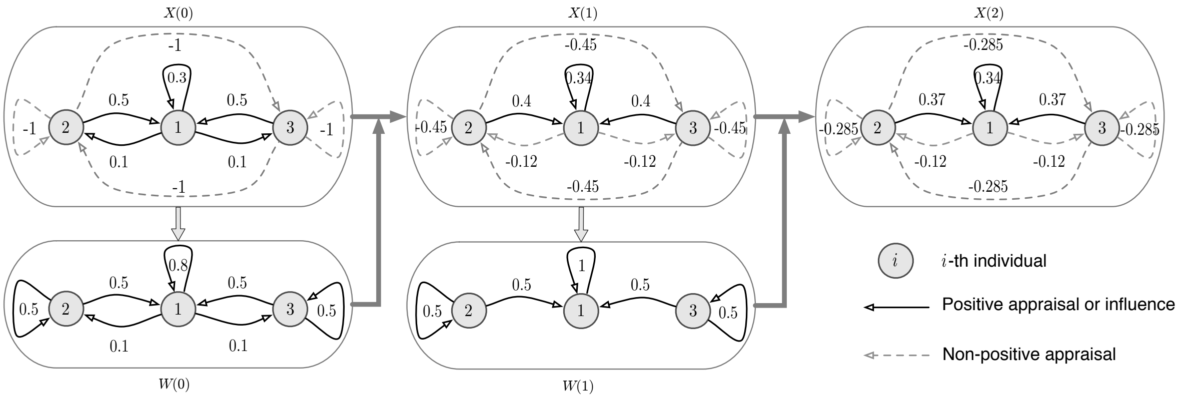


Fig. 2. Appraisal and influence coevolution in a triad: the digraphs formulated by solid directed edges correspond to positive appraisal digraphs (on the top) and influence digraphs (on the bottom). The non-positive appraisals are also shown by dotted lines. The numbers on the edges are the values of the appraisals and influence weights, respectively. As stated in Theorem 4.1, regarding the positive appraisal digraphs, (1) the (positive) edges from individual 1 to the individuals 2 and 3 disappear at time $t = 1$ and these edges will never appear again; (2) the edges between the individuals 2 and 3 never appear. Hence, three SCCs remain unchanged after $t = 1$.

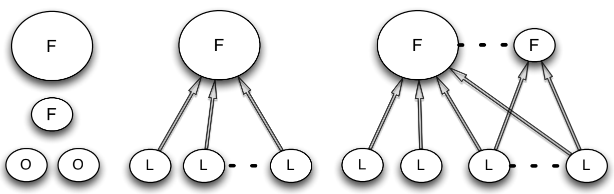


Fig. 3. Faction-follower-outsiders structure: "F" components are factions, "L" components are followers, "O" components are outsiders. The small size components are singletons and the large size components are SCCs with two or more nodes. On the left: two factions and two outsiders. In the middle: a factions-followers structure with a single faction. On the right: a factions-followers structure with multiple factions.

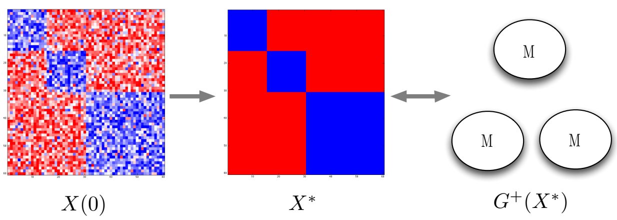


Fig. 4. Convergence to three factions.

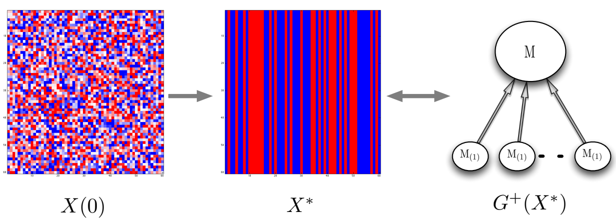


Fig. 5. Convergence to a factions-followers structure with a single faction.

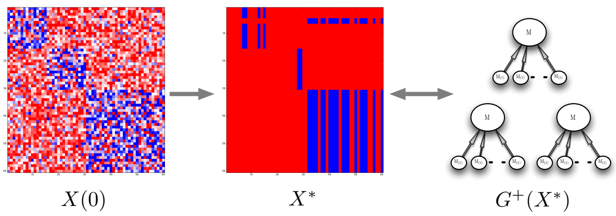


Fig. 6. Convergence to a factions-followers structure: three disconnected factions-followers structures each of which has a single faction.

The positive digraph could be either reducible or irreducible. Moreover, in the equilibrium, such a structure has only one faction, an arbitrary number of followers and no outsider. Regarding the irreducibility assumption of the third statement of Theorem 5.2, we can show that if $G^+(X(0))$ is irreducible, then $G^+(X(t))$ is irreducible for all $t \geq 0$ and $G^+(\lim_{t\to\infty} X(t))$ is irreducible. This statement can be extended such that a sink SCC (with at least two nodes) becomes a faction in the equilibrium.

5.2 Numerical study on empirical networks

We now apply our results to empirical social network examples. To minimize the impact of the self-appraisal constant ϵ on the trajectory of the coevolution system, we select small values of ϵ in the simulations. We report trajectories computed for strictly-positive small ϵ, but we comment that essentially identical trajectories are generated by setting ϵ to zero.

In our first example, we consider the appraisal evolution on a Krackhardt's advice network [34]; this network describes a manufacturing organization with 21 managers and 128 relationships in which a manager sought advice from another manager. Because the available data about this advice network does not include a complete set of interpersonal appraisals, we set up an initial appraisal matrix based on two ancillary assumptions: (1) each individual holds initial interpersonal appraisals equal to 1 of the individuals she seeks advice from and has non-positive appraisals (uniformly randomly-selected from $[-0.5, 0]$) of all other individuals; (2) the initial self-appraisals are equal to the normalized in-degree of the individuals in the advice network. Note that the ancillary assumption 2) is grounded in the theory of reflected appraisal as presented in [35]. Given $\epsilon = 0.1$, the trajectory of the appraisal evolution is shown in Fig. 7. The equilibrium structure is a factions-followers structure with a single faction. Appraisal consensus of each individual in the network is observed. Note that this convergence to an appraisal structure with a single faction is robust with respect to the initial non-positive appraisal assignment among $[-0.5, 0]$. Indeed, our initial appraisal assignment guarantees that at least two (the 2nd and the 21st) columns of $X(1)$ are positive. Based on statement (ii) in Theorem 5.2, there always exists a globally reachable node in the equilibrium structure. That is to say, the equilibrium must have a factions-followers structure with a single faction, with or without followers. However, this claim is not true for an arbitrary initial appraisal assignment.

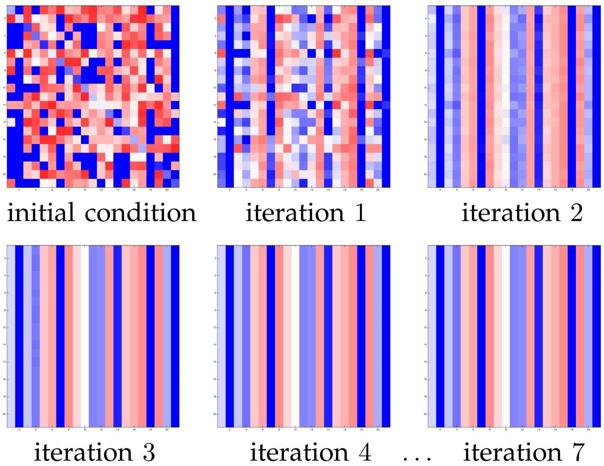


Fig. 7. Evolution of appraisal matrices on Krackhardt's advice network: the interpersonal appraisals converge to a rank-1 matrix of the form $\mathbb{1}_N v^T$.

The second example considers the social interactions among a group of monks in an isolated contemporary American monastery observed by Sampson [36]. Based on observations and experiments, Sampson collected a variety of experimental information on four types of relations: Affect, Esteem, Influence, and Sanctioning. Each of 18 respondent monks ranked their three first choices on these relations, where 3 indicates the highest or first choice and 1 the last choice in the presented interaction matrices. Some subjects offered tied ranks for their top five choices. Here we focus on the monastery appraisal structures based on the ranking of the affection ("like" and "dislike") relations in Sampson's empirical data. Note that we apply data collected directly from original PhD dissertation [36], where "like" and "dislike" relations were both collected for three times, while most Sampson's "dislike" dataset available online are incomplete.

Because the available data about this network does not include a complete set of interpersonal appraisals, we set up an initial appraisal matrix for our simulation based on one ancillary assumption: the initial self-appraisal of each individual is equal to the mean value of the appraisals of this individual held by all other group members. Given $\epsilon = 0.1$, the trajectories of the appraisal evolution on Sampson's affection network measured for the third time are shown in Fig. 8. The equilibrium structure illustrated in Fig. 8 is still

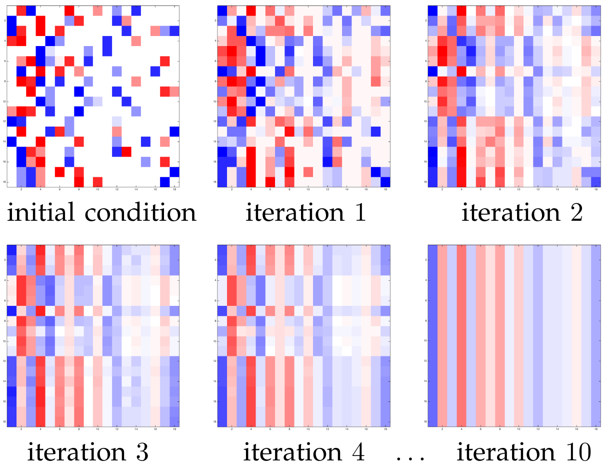


Fig. 8. Evolution of appraisal matrices on Sampson's monastery network: the appraisal matrix converges to a rank-1 matrix of the form $\mathbb{1}_N v^T$.

a factions-followers structure with a single faction. Appraisal consensus of each individual in the network is also observed along the trajectories. Moreover, we observe a quick consolidation to approximately two clusters (two factions-followers structures) on a short time-scale, but then on a long time-scale, the bridging ties with positive appraisals and influences bring the whole group together slowly, and eventually, one faction with followers emerges in the equilibrium.

The third example considers Zachary's karate club network. The interactions among the members of a university karate club were recorded for 2 years by Zachary [37]. During observation, a conflict between the administrator and the instructor of the club developed and eventually the club broke into two clubs.

Because the available data about this network does not include a complete set of interpersonal appraisals, we set up an initial appraisal matrix for our simulation based on four ancillary assumptions: (1) each individual in the group has positive initial appraisals of the individuals that she interacted with and the appraisal values are proportional to the number of contexts in which interaction took place between the two individuals; (2) each individual has non-positive appraisals of the remaining individuals and the appraisal values are uniformly randomly-selected from $[-0.5, 0]$, while the administrator and the instructor have -1 appraisal of each other; (3) the initial self-appraisal of each individual is equal to the mean value of the positive appraisals of this individual held by the others; moreover, (4) the self-influence parameter for the two intransigent individuals (i.e., the administrator and the instructor) is $\epsilon = 1$, and $\epsilon = 0$ for the remaining individuals. The trajectory of the appraisal evolution on Zachary's karate club network is shown in Fig. 9. Consistent with Zachary's analysis, we observe two factions-followers structures emerged in the equilibrium and each structure has a single faction consisting of either the instructor (node 1) or the administrator (node 34).

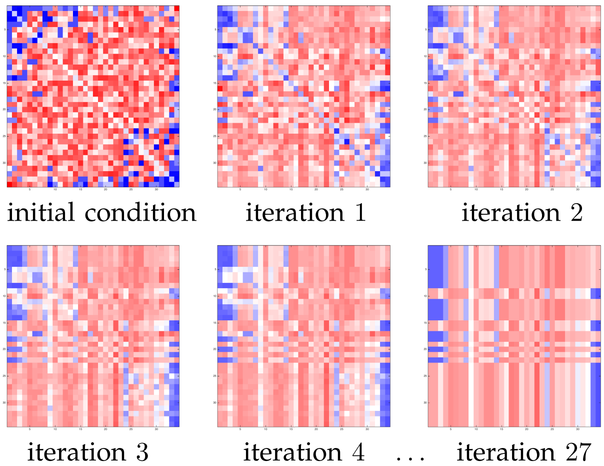


Fig. 9. Evolution of appraisal matrices on Karate club network: the equilibrium positive digraph includes two factions-followers structures. Node 1 corresponds to the instructor and Node 34 corresponds to the administrator. The appraisal submatrices associated with two structures converge to two rank-1 matrices.

5.3 Theoretical and numerical analysis of the self-influence parameter

To complete the analysis, we discuss the self-influence parameter ϵ and show its impact on the appraisal and influence coevolution. In particular, for $\epsilon = 0$, if we additionally assume that the influence (and equivalently positive appraisal) submatrices associated with the sink SCCs are aperiodic for all time, then all results in Theorems 5.1 and 5.2 hold true. The proofs are similar and skipped here. Moreover, even without an aperiodicity assumption, given $\epsilon = 0$, $X_s(t)$ in the statement (i) of Theorem 5.2 still exponentially converges to a rank-1 positive matrix of the form $\mathbb{1}_{N_s} v^T$, and consequentially the statement (ii) holds. That is, aperiodicity is implicitly satisfied for these two cases if there exists a positive column for the considered appraisal submatrix $X_s(t)$ or matrix $X(t)$. In addition, ϵ does not need to be homogeneous for each individual and one may verify that all results in this article hold for positive and heterogeneous $\{\epsilon_i\}_{i \in \{1,\dots,N\}}$. Heterogeneous self-appraisal constants are adopted in the simulation of Fig. 9.

By the following numerical simulations, we claim that the equilibrium appraisal structure and the convergence rate may vary for different ϵ. Consider a coevolution system with 10 individuals. Given a constant initial state, we show the dynamical trajectories for three different ϵ. For $\epsilon = 0$, the dynamical system converges to an $O(10^{-5})$-neighborhood of the equilibrium in 7 iterations, and the topology evolutions of the positive appraisal digraphs and their condensation digraphs are shown in Fig. 10. For $\epsilon = 0.5$, the system converges in 28 iterations to an $O(10^{-5})$-neighborhood of the equilibrium as shown in Fig. 11. For $\epsilon = 0.9$, the topology evolutions of the digraphs are referred to Fig. 12. It takes 41 iterations in this case to reach an $O(10^{-5})$-neighborhood of the equilibrium. The simulations illustrate that a larger ϵ essentially corresponds to a slower convergence rate. It is easy to understand as ϵ represents the self-influence parameter of individuals, which measures the stubbornness of each individual on its previous opinion. From Figures 10–12, we also observe different trajectories of the appraisal structure evolutions for different ϵ.

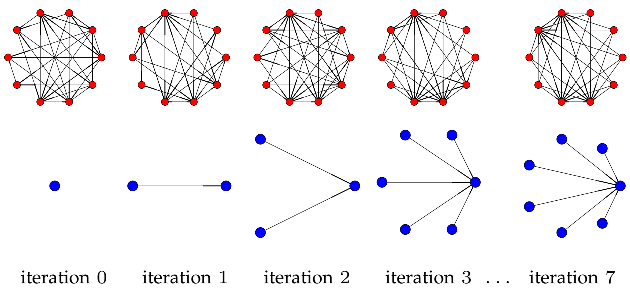


Fig. 10. Topology evolution of an appraisal structure with $\epsilon = 0$: in this and following three figures, the digraphs above (resp. below) correspond to the evolution of $G^+(X(t))$ (resp. $C(G^+(X(t)))$). The equilibrium positive digraph includes one faction (consisting of 4 nodes) and six followers.

Moreover, the number of factions at equilibrium may also vary for different ϵ. As illustrated in Fig. 13, given a coevolution system with 10 individuals and a constant initial state, the equilibrium appraisal structure has one factions-followers structure with a single faction for $\epsilon = 0.1$ and has two disconnected factions-followers structures each of which has a single faction for $\epsilon = 0.9$.

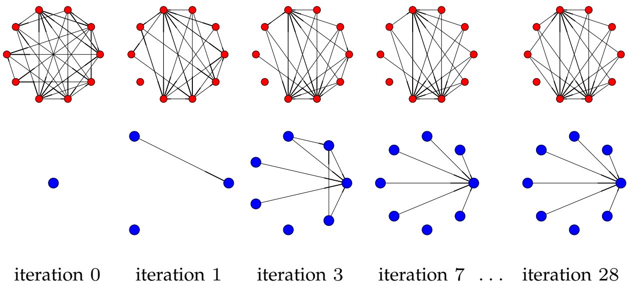


Fig. 11. Topology evolution of an appraisal structure with $\epsilon = 0.5$. The equilibrium positive digraph has two disconnected components. The component with 9 nodes has one faction and six followers. Another disconnected component includes an outsider.

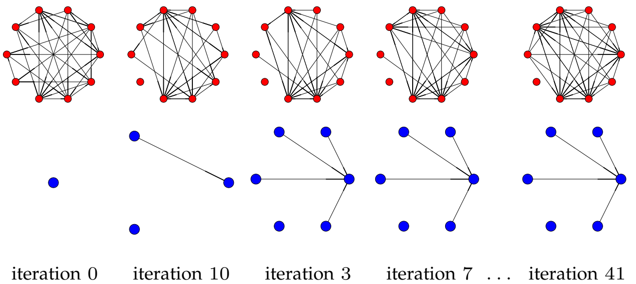


Fig. 12. Topology evolution of an appraisal structure with $\epsilon = 0.9$. The equilibrium positive digraph has one outsider and one factions-followers component. The factions-followers component includes 9 nodes: one faction (consisting of 5 nodes) and four followers.

6 Structural Balance Properties of the Coevolutionary Dynamics

In this section we study the structural balance properties of the coevolution model of interpersonal appraisal and influence (4)-(5). In the previous two sections we have illustrated the topology evolution for the positive appraisal digraphs and the convergence properties for the appraisal and influence coevolution. Now we are able to combine these results with macro-structural models: it is interesting to study what macro-structural models the equilibrium factions-followers-outsiders structure is related to and which class of macro-structures are invariant under the coevolution. In what follows the word macro-structure is a synonym for an appraisal structure, i.e., a set of all appraisal matrices with a certain sign pattern. A macro-structure is *invariant* under the coevolution if, given an initial appraisal matrix belonging to the macro-structure, all trajectory matrices via the coevolution system (4)-(5) remain in the macro-structure.

Our coevolution model approach does not pre-specify a particular micro-model. Instead, it pre-specifies the conditions of interpersonal influence relations and addresses the implications of the model.

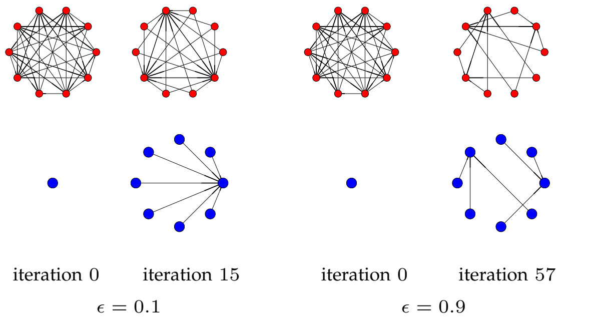


Fig. 13. Topology evolutions of an appraisal structure with $\epsilon = 0.1$ and $\epsilon = 0.9$, respectively. The equilibrium positive digraph has one faction for $\epsilon = 0.1$ but two factions for $\epsilon = 0.9$.

Recall that our coevolution model satisfies the two statements in the classic balance model: "my friend's enemy is my enemy" and "my friend's friend is my friend", whereas the other two statements: "my enemy's enemy is my friend" and "my enemy's friend is my enemy" are not intuitively necessary for the coevolution of appraisal and influence. By examining all 16 types of triads in an appraisal structure, the deduced micro-model of permitted triad types by the first two statements is $\{300, 120D, 102, 021U, 012, 003\}$ (see Fig. 14 and [8] and [27, Section 8.3] for the detailed description of these triad types). Moreover, as we allow the interpersonal appraisal relation to be 0, the triad type $021D$ is also permitted in our model if the two bottom nodes of the positive digraph of $021D$ in Fig. 14 have 0 appraisal of each other. Overall, the micro-model associated with the coevolution model (4)-(5) is $P_{\text{co-evolv}} = \{300, 120D, 102, 021D, 021U, 012, 003\}$.

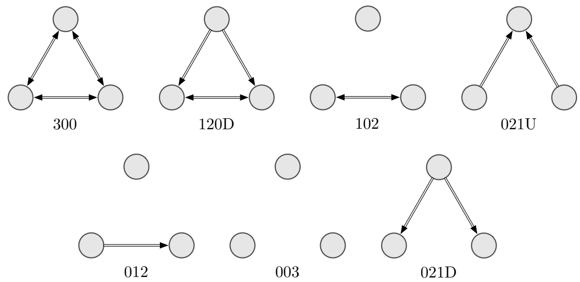


Fig. 14. Permitted triads: the positive appraisal digraph representations of the permitted triad types by the coevolution model

Furthermore, we examine the equilibrium structures described in Theorem 5.1. It is clear that the factions-followers-outsiders structure is the macro-structure defined by the micro-model $P_{\text{co-evolv}}$, where only triad types in $P_{\text{co-evolv}}$ appear in the structure and all remaining triad types are forbidden. In particular, triad type 300 is a one-faction structure, $120D$ is a one-faction-one-follower structure, 102 is a one-faction-one-outsider structure or a two-faction structure (depending on the top individual's self-appraisal), $021D$ is a two-faction-one-follower structure (where the interpersonal appraisals between the factions are 0), $021U$ is a one-faction-two-follower structure, 012 is a one-faction-one-follower-one outsider/faction structure, and finally 003 includes three factions or outsiders. Consequently, any appraisal network including only triad types in $P_{\text{co-evolv}}$ has a factions-followers-outsiders structure.

Proposition 6.1. *The factions-followers-outsiders structure, i.e., the equilibrium macro-structure of the coevolution system* (4)-(5)*, is consistent with the micro-model* $P_{\text{co-evolv}}$.

In other words, the coevolution system bridges the static micro-model and the dynamical convergence of the macro-structure networks, and the factions-followers-outsiders networks are then structurally balanced with respect to the micro-model $P_{\text{co-evolv}}$.

6.1 Invariant macro-structures

We have studied the macro-structure associated with the coevolution equilibrium appraisal networks. In the following, we analyze the macro-structures which are invariant in the coevolution system (4)-(5).

The implication of the *classical model of structural balance* is a class of appraisal macro-structures where either all individuals have strictly positive appraisal relations, or there are at most two subgroups such that individuals have strictly positive appraisal relations in the same subgroup but strictly negative appraisal relations between two subgroups. A classical balance structure has two possible block matrix patterns: (1) $[+]$ or (2) $\begin{bmatrix} D_1 & - \\ - & D_2 \end{bmatrix}$, where $\{D_i\}_{i\in\{1,\dots,2\}}$ are M-cliques. Moreover, $D_i \sim [+]$ if there are $N_i \geq 2$ individuals in this M-clique, and D_i can be either "$-$" or "$+$" for $N_i = 1$. It is noted that "M-clique" and "faction" are two similar but different concepts in this paper: the differences lie on (1) an outsider is a stand-alone M-clique but not a faction, (2) there may exist A-relations between two M-cliques but never between two factions. The classical balance structure is a special case of a factions-followers-outsiders structure, which includes at most two factions, may include outsiders, but does not include any followers. The classical balance macro-structure has been intensively studied, see e.g., in [1], [2]. We also consider other two macro-structural models: *clustering structure* and *ranked clusters of M-clique structure* in the following analysis.

Lemma 6.2 (Invariance of classical balanced structure)**.** *The classical balanced structure is invariant under the coevolution system* (4)-(5).

Define a *clustering structure* as an appraisal structure with a representative matrix

$$X \sim \begin{bmatrix} D_1 & - & \dots & - \\ - & D_2 & \dots & - \\ \vdots & \vdots & \ddots & \vdots \\ - & - & \dots & D_n \end{bmatrix}.$$

Here D_i, $i \in \{1,\dots,n\}$ are M-cliques (clusters), and all "$-$" block submatrices represent complete N-relations among these M-cliques. That is to say, the clustering structure extends the classical balanced structure to a structure with $n > 2$ M-cliques. This structure is also a factions-followers-outsiders structure, with an arbitrary number of factions.

Lemma 6.3 (Invariance of clustering structure)**.** *The clustering structure is invariant under the coevolution system* (4)-(5).

A *ranked clusters of M-clique structure* is defined by a block matrix form: $X \sim \begin{bmatrix} D_1 & - & \dots & - \\ S_{21} & D_2 & \dots & - \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \dots & D_n \end{bmatrix}$, with n M-cliques for $n \geq 2$. Here, without loss of generality, if $i \geq j$, the rank of the i-th M-clique is higher than or equal to the rank of the j-th M-clique. S_{ij} is strictly positive if and only if the i-th M-clique ranks strictly higher than the j-th M-clique; otherwise, if the i-th and the j-th M-cliques have the same rank, then S_{ij} is non-positive. One may check the ranked clusters of M-clique structure is not invariant under the coevolution system (4)-(5) in general. However, we will show that a subset of this structure is invariant under the coevolution.

Lemma 6.4 (Invariance of ranked clusters of M-clique structure)**.** *A ranked clusters of M-clique structure is not invariant under the coevolution system* (4)-(5) *in general. But, if an appraisal matrix X has both a ranked clusters of M-clique structure and a factions-followers structure with only one faction, then the structure of X is invariant under the coevolution.*

One may check that X in Lemma 6.4 satisfies

$$X \sim \begin{bmatrix} + & \dots & + & - & \dots & - \\ + & \dots & + & - & \dots & - \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ + & \dots & + & - & \dots & - \end{bmatrix}, \tag{8}$$

that is, all entries in the same columns of X have the same sign. It is noted that the structure of X has totally two ranks: only one M-clique is with the higher rank and it is a faction, and the remaining $n-1$ M-cliques have the same lower rank and they are followers, as shown in Fig. 15.

Among all macro-structures introduced in [27, Section 8.3], the three classes of macro-structures discussed in Lemmas 6.2–6.4 are all potentially stable balanced structures under our coevolution. It is noted that the equilibrium appraisal structure also includes

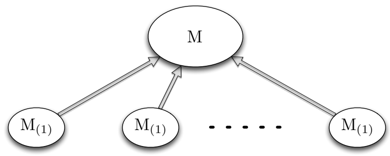


Fig. 15. Positive digraph of the invariant ranked clusters of M-clique structure: The nodes of the graph are M-cliques. The top M-clique is a faction and all $M_{(1)}$ are followers. The structure is a factions-followers structure with a single faction.

a ranked clusters of M-cliques structure specified as in Fig. 16. Different from the structure in Fig. 15, the structure in Fig. 16 has multiple factions and each follower may hold positive appraisals of more than one faction. One simple example for the equilibrium appraisal matrix in this case is that $X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1/2 & 1/2 & 0 \end{bmatrix}$ for any $\epsilon \geq 0$. However, this structure is not invariant under the coevolution system in general. Moreover, such an equilibrium is less-frequently observed in simulations with random initial conditions. In the example above, if the appraisal of node 1 held by the follower, node 3, increases for a sufficiently small amount, then the trajectory of the coevolution system leads to another equilibrium where node 3 is only directly connected to node 1.

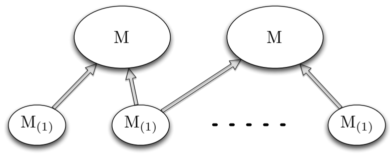


Fig. 16. Positive digraph of another equilibrium ranked clusters of M-clique structure: each sink M-clique is a faction; all source singleton $M_{(1)}$, are followers that hold (complete) positive appraisals of one or multiple factions. The structure is a factions-followers structure with multiple factions.

6.2 Convergence of invariant macro-structures

Now we integrate the invariant structure results in Lemmas 6.2–6.4 with the convergence results in Section 5, which immediately implies the convergence properties of the stable macro-structures as in Corollary 6.1. In what follows we regard the classical balanced structure with two M-cliques as a special case of a clustering structure for the simplicity of presentation.

Corollary 6.1 (Convergence of generalized balanced structures). *For the coevolution system* (4)-(5) *with* $\epsilon \geq 0$*, each trajectory* $X(t)$ *converges exponentially fast to an equilibrium* X^* *in the following three scenarios:*

(i) (Convergence of a classical balanced structure with one cluster) *For a group of individuals with positive initial appraisals,* $G^+(X(t))$ *is a faction for all* $t \geq 0$ *and so is* $G^+(X^*)$*. Moreover, a positive appraisal consensus on each individual is achieved for the whole group in* X^**.*

(ii) (Convergence of a clustering structure) *For a group of individuals with a clustering appraisal structure initially, the factions and outsiders of* $G^+(X(0))$ *remain unchanged in* $G^+(X(t))$ *for all* $t \geq 0$ *and in* $G^+(X^*)$*. An appraisal consensus of each individual of the group is achieved within each faction of* $G^+(X^*)$*: it is positive if the individual belongs to the faction and non-positive otherwise. An outsider occurs if and only if one cluster includes one individual and its self-appraisal is non-positive.*

(iii) (Convergence of a ranked clusters of M-clique structure with form (8)) *For a group of individuals with an initial appraisal structure* (8)*, the factions-followers structure with one faction remains unchanged in* $G^+(X(t))$ *for all* $t \geq 0$ *and in* $G^+(X^*)$*. The signs of all appraisals never change along the trajectory* $X(t), t \geq 0$*, and an appraisal consensus on each individual is achieved for the whole group in* X^**.*

Different from Theorem 5.2 (iii), the first statement (i) of Corollary 6.1 assumes that all appraisals of the initial state are strictly positive, which implies the aperiodicity and irreducibility of all $X(t)$ and $W(t)$ along the trajectory. Therefore, ϵ could be equal to 0. Similarly, ϵ could be 0 for the second statement (ii). The third statement (iii) is a special case of Theorem 5.2 (ii), and therefore, the aperiodicity is satisfied implicitly and the statement holds for $\epsilon = 0$.

7 Conclusion

This article studies appraisal structure evolution among a group of individuals. Motivated by recent efforts on developing linkages between the major topics in sociological social psychology, we believe that it is interesting and meaningful to link social influence network theory with structural balance theory. As appraisals are subject to endogenous interpersonal influences, they may be influenced by others' appraisals. A network of such endogenous interpersonal influences is often formed in social groups. However, to the best of our knowledge, there are no dynamical models of appraisal structure which are directly evolved with the implications of such influence networks. It is not theoretically clear how the fundamental appraisals associated with persons' social identities are modified by the displayed influences of other group members, or how endogenous interpersonal influences in a group may generate equilibrium appraisals that are quite different from the initial array of appraisals.

We have presented novel results on the modeling and analysis of the coevolution of appraisal and influence networks. We derived a concise explicit dynamical model for the coevolution process and characterized completely its convergence and equilibrium structure properties. Our analysis also leads to several important implications to the study of signed social networks and structural balance theory. Specifically, our model shows that (i) for any initial appraisal matrix, the set of strongly connected components associated with the positive appraisal digraphs remains constant after finite time; (ii) for any initial appraisal matrix, the appraisal matrix trajectory converges asymptotically to an equilibrium, which has a factions-followers-outsiders structure: all individuals in a faction reach an appraisal consensus on each individual, all followers' appraisals are determined by the appraisals held by the individuals from the directly connected factions, and all outsiders have non-positive appraisals of each individual; and (iii) the appraisal structures according to the equilibria of the coevolution are balanced in sense that the two statements "my friend's enemy is my enemy" and "my friend's friend is my friend" are always satisfied in the associated social networks. The realizations of all possible equilibria of the coevolution fall into four distinct social structural classes. Meanwhile, three macro structural models are proved to be always stable subject to the proposed coevolution process. Overall, our model predicts a tendency of social appraisal structures to a set of structural equivalent bundles, i.e., a set of components where individuals have aligned interpersonal appraisals.

This paper presents only an introduction to appraisal evolution and structural balance models with implications of social influence networks, and much work remains to be done in order to understand the robustness of our formulation and its results. We assume here that the influence weights accorded by each individual are proportional to her positive appraisals on individuals of the social group. However, a large literature exists in social psychology on conditions that may affect individuals' influence network and its evolution (e.g., see our recent work [20]). We believe there are opportunities for a discussion on useful alternative mechanisms that adjust the relation between interpersonal appraisals and influences. Future research will be directed at validating our results with empirical data and identifying the qualitative roles of appraisal and influence coevolution mechanisms in the dynamics of signed social networks.

Acknowledgments

The authors are grateful to Professors Alex Olshevsky and Wei Ren for their comments on Lemma B.1. They also thank the editors and anonymous reviewers for their excellent comments and suggestions.

References

[1] S. A. Marvel, J. Kleinberg, R. D. Kleinberg, and S. H. Strogatz, "Continuous-time model of structural balance," *Proceedings of the National Academy of Sciences*, vol. 108, no. 5, pp. 1771–1776, 2011.

[2] G. Facchetti, G. Iacono, and C. Altafini, "Computing global structural balance in large-scale signed social networks," *Proceedings of the National Academy of Sciences*, vol. 108, no. 52, pp. 20 953–20 958, 2011.

[3] V. A. Traag, P. V. Dooren, and P. D. Leenheer, "Dynamical models explaining social balance and evolution of cooperation," *PLOS ONE*, vol. 8, no. 4, p. e60063, 2013.

[4] D. Cartwright and F. Harary, "Structural balance: a generalization of Heider's theory," *Psychological Review*, vol. 63, no. 5, p. 277, 1956.

[5] F. Heider, "Attitudes and cognitive organization," *The Journal of Psychology*, vol. 21, no. 1, pp. 107–112, 1946.

[6] ——, *The Psychology of Interpersonal Relations*. Lawrence Erlbaum Associates, 1958.

[7] J. A. Davis and S. Leinhardt, "The structure of positive interpersonal relations in small groups," in *Sociological Theories in Progress*, J. Berger, M. Zelditch, and B. Anderson, Eds. Houghton-Mifflin, 1972, vol. 2, pp. 218–251.

[8] E. C. Johnsen, "The micro-macro connection: Exact structure and process," in *Applications of Combinatorics and Graph Theory to the Biological and Social Sciences*, F. Roberts, Ed. Springer, 1989, pp. 169–201.

[9] K. Kułakowski, P. Gawroński, and P. Gronek, "The Heider balance: a continuous approach," *International Journal of Modern Physics C*, vol. 16, no. 05, pp. 707–716, 2005.

[10] E. C. Johnsen, "Network macrostructure models for the Davis-Leinhardt set of empirical sociomatrices," *Social Networks*, vol. 7, no. 3, pp. 203–224, 1985.

[11] T. Antal, P. L. Krapivsky, and S. Redner, "Dynamics of social balance on networks," *Physical Review E*, vol. 72, no. 3, p. 036121, 2005.

[12] S. A. Marvel, S. H. Strogatz, and J. M. Kleinberg, "Energy landscape of social balance," *Physical Review Letters*, vol. 103, p. 198701, 2009.

[13] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical Association*, vol. 69, no. 345, pp. 118–121, 1974.

[14] N. E. Friedkin and C. E. Johnsen, "Attitude change, affect control, and expectation states in the formation of influence networks," *Advances in Group Processes*, vol. 20, pp. 1–29, 2003.

[15] D. Snyder and E. L. Kick, "Structural position in the world system and economic growth, 1955-1970: A multiple-network analysis of transnational interactions," *American Journal of Sociology*, pp. 1096–1126, 1979.

[16] P. R. Krugman, *The Self Organizing Economy*. Blackwell Publishers, 1996.

[17] E. O. Laumann and F. U. Pappi, *Networks of collective action: A perspective on community influence systems*, ser. Quantitative Studies in Social Relations. Academic Press, 1976.

[18] S. P. Borgatti and M. G. Everett, "Models of core/periphery structures," *Social Networks*, vol. 21, no. 4, pp. 375–395, 2000.

[19] T. Gross and B. Blasius, "Adaptive coevolutionary networks: A review," *Journal of The Royal Society Interface*, vol. 5, no. 20, pp. 259–271, 2008.

[20] P. Jia, A. MirTabatabaei, N. E. Friedkin, and F. Bullo, "Opinion dynamics and the evolution of social power in influence networks," *SIAM Review*, vol. 57, no. 3, pp. 367–397, 2015.

[21] M. O. Jackson, "A survey of models of network formation: Stability and efficiency," in *Group Formation in Economics; Networks, Clubs and Coalitions*, G. Demange and M. Wooders, Eds. Cambridge University Press, 2005.

[22] B. Skyrms and R. Pemantle, "A dynamic model of social network formation," *Proceedings of the National Academy of Sciences*, vol. 97, no. 16, pp. 9340–9346, 2000.

[23] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[24] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

[25] A. B. Kahn, "Topological sorting of large networks," *Communications of the ACM*, vol. 5, no. 11, pp. 558–562, 1962.

[26] P. Bremaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, 1999.

[27] N. E. Friedkin, *A Structural Theory of Social Influence*, ser. Structural Analysis in the Social Sciences. Cambridge University Press, 1998.

[28] J. A. Bargh and M. J. Ferguson, "Beyond behaviorism: on the automaticity of higher mental processes," *Psychological Bulletin*, vol. 126, no. 6, pp. 925–945, 2000.

[29] D. Kahneman, *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.

[30] I. Ajzen, "Nature and operation of attitudes," *Annual Review of Psychology*, vol. 52, no. 1, pp. 27–58, 2001.

[31] J. R. P. French, "A formal theory of social power," *Psychological Review*, vol. 63, no. 3, pp. 181–194, 1956.

[32] F. Harary, "A criterion for unanimity in French's theory of social power," in *Studies in Social Power*, D. Cartwright, Ed. University of Michigan, 1959, pp. 168–182.

[33] N. H. Anderson, *Foundations of Information Integration Theory*. Academic Press, 1981.

[34] D. Krackhardt, "Cognitive social structures," *Social Networks*, vol. 9, no. 2, pp. 109–134, 1987.

[35] N. E. Friedkin, "A formal theory of reflected appraisals in the evolution of power," *Administrative Science Quarterly*, vol. 56, no. 4, pp. 501–529, 2011.

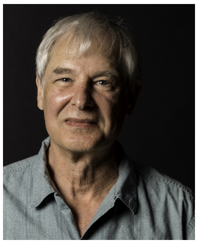
[36] S. F. Sampson, "Crisis in a cloister," Ph.D. dissertation, Cornell University, 1969.

[37] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977.

[38] L. Moreau, "Stability of multiagent systems with time-dependent communication links," *IEEE Transactions on Automatic Control*, vol. 50, no. 2, pp. 169–182, 2005.

[39] J. M. Hendrickx, "Graphs and networks for the analysis of autonomous agent systems," Ph.D. dissertation, Université Catholique de Louvain, Belgium, Feb. 2008.

Peng Jia is an assistant project scientist in the Department of Mechanical Engineering at the University of California, Santa Barbara. He received a Ph.D. degree in Electrical and Computer Engineering from McGill University, Montreal, Canada, in 2010, a M.S. degree from the Chinese University of Hong Kong, Hong Kong, in 2005, and a B.S. degree from Beihang University, Beijing, China, in 2001. His research interests include opinion dynamics and social networks, auctions and their applications in decentralized systems, stochastic systems and control, stochastic games, and nonlinear control and output regulation.



Noah E. Friedkin is a Professor in the Sociology Department and the Center for Control, Dynamical Systems and Computation at the University of California, Santa Barbara, with a B.A and Ph.D. from the University of Chicago. His areas of research are social psychology (social networks and group processes related to opinion dynamics), mathematical sociology, and complex social organizations. He is author of A Structural Theory of Social Influence (Cambridge, 1998) and co-author of Social Influence Network Theory: A Sociological Examination of Small Group Dynamics (Cambridge, 2011) both of which received Outstanding Book Awards from the Mathematical Sociology Section of the American Sociological Association. His article publications have appeared in numerous journals, including the Administrative Science Quarterly, American Journal of Sociology, American Sociological Review, Social Psychology Quarterly, IEEE Control Systems, and SIAM Review. He is Editor in Chief of the Journal Mathematical Sociology.

Francesco Bullo is a Professor with the Mechanical Engineering Department and the Center for Control, Dynamical Systems and Computation at the University of California, Santa Barbara. He was previously associated with the University of Padova, the California Institute of Technology, and the University of Illinois at Urbana-Champaign. His research interests focus on network systems and distributed control with application to robotic coordination, power grids and social networks. He is the coauthor of "Geometric Control of Mechanical Systems" (Springer, 2004) and "Distributed Control of Robotic Networks" (Princeton, 2009). His articles received the 2008 CSM Outstanding Paper Award from IEEE CSS, the 2011 Hugo Schuck Best Paper Award from AACC, the 2013 SIAG/CST Best Paper Prize from SIAM, the 2014 Automatica Best Paper Prize from IFAC, and the 2016 Guillemin-Cauer Best Paper Award from IEEE CAS. He is a Fellow of IEEE and IFAC. He has served on the editorial boards of IEEE, SIAM, and ESAIM journals, and on the Executive Committee of the IEEE Control Systems Society.

APPENDIX A
PROOF OF THEOREM 4.1

From the definition, any digraph associated to a non-negative matrix $X^+(t) \geq 0$ has at least one SCC. As some entries of $X(t)$ may be negative, it is possible for one SCC of $G^+(X(t))$ to split into multiple SCCs at time $t+1$. For example, given

$$X(t) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ with } a, b, c > 0 \text{ and } d \leq 0,$$

by simple calculation,

$$W(t) = \begin{bmatrix} \frac{a+\epsilon}{a+\epsilon+b} & \frac{b}{a+\epsilon+b} \\ \frac{c}{c+\epsilon} & \frac{\epsilon}{c+\epsilon} \end{bmatrix}, \quad \text{and}$$

$$X(t+1) \sim \begin{bmatrix} + & \frac{b(a+\epsilon)}{a+\epsilon+b} + \frac{bd}{a+\epsilon+b} \\ + & \frac{bc}{c+\epsilon} + \frac{\epsilon d}{c+\epsilon} \end{bmatrix}.$$

It is clear that there is only one SCC for $X^+(t)$ but there are two SCCs for $X^+(t+1)$ if $|d| > a + \epsilon$. The result can be extended to any dimension matrix $X(t)$.

Regarding facts (i)-(ii), we first consider the case where the positive digraph at time t includes two SCCs. After performing an inverse topological sorting, we have a block lower triangular representative of the influence matrix as follows:

$$W(t) = \begin{bmatrix} D_1 & 0 \\ S_1 & D_2 \end{bmatrix}, \tag{9}$$

where the block submatrices $D_1 \in \mathbb{R}^{N_1 \times N_1}, D_2 \in \mathbb{R}^{(N-N_1)\times(N-N_1)}$, correspond to the SCCs $H_1(t)$ and $H_2(t)$ respectively, and $D_1, S_1, D_2 \geq 0$. By definition, $X(t)$ shall have the form

$$X(t) = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, \text{ with } X_{12} \leq 0.$$

Clearly,

$$X^+(t) = \begin{bmatrix} X_{11}^+ & 0 \\ X_{21}^+ & X_{22}^+ \end{bmatrix},$$

and

$$X(t+1) = \begin{bmatrix} D_1X_{11} & D_1X_{12} \\ S_1X_{11} + D_2X_{21} & S_1X_{12} + D_2X_{22} \end{bmatrix},$$

where $DX_{12} \leq 0$ for $X_{12} \leq 0$. As a result,

$$X^+(t+1) = \begin{bmatrix} (D_1X_{11})^+ & 0 \\ (S_1X_{11} + D_2X_{21})^+ & (S_1X_{12} + D_2X_{22})^+ \end{bmatrix}. \tag{10}$$

By comparing $X^+(t+1)$ with $X^+(t)$, we have the following conclusion: if no node of the SCC $H_1(t)$ has an influence link toward a node of the SCC $H_2(t)$ or equivalently no individual in $H_1(t)$ likes an individual in $H_2(t)$ at time t, then no edge can appear from $H_1(t)$ to $H_2(t)$ in the positive digraph $G^+(X(t+1))$ at time $t+1$. Moreover, the subset of nodes $H_1(t)$ (or $H_2(t)$ respectively) either remains one SCC, or may split into multiple SCCs, depending on the structure of $[D_1X_{11}]^+$ (or $[S_1X_{12} + D_2X_{22}]^+$ respectively) in the digraph $G^+(X(t+1))$. It is impossible for the cardinality of an SCC to increase.

These results can be extended to any higher dimension case. Consider a block lower triangular influence matrix

$$W(t) = \begin{bmatrix} D_1(t) & & & & & \\ \vdots & \ddots & & & & \\ S_{i1}(t) & \dots & D_i(t) & & \mathbf{0} & \\ \vdots & \ddots & \vdots & \ddots & & \\ S_{j1}(t) & \dots & S_{ji}(t) & \dots & D_j(t) & \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \end{bmatrix}. \tag{11}$$

The diagonal block matrices $D_i(t)$ and $D_j(t)$ correspond to the nodes $H_i(t), H_j(t)$ of the associated condensation digraph at time t respectively and $i < j$ in the inverse topological sorting. By the similar analysis above, we can show that all zero entries in the upper triangular will remain zero for $W(t+1)$ and $X^+(t+1)$ via the coevolution system (4)-(5). That is, no edge exists from the subset of nodes $H_i(t)$ to the subset of nodes $H_j(t)$ at time $t+1$ and hence no two SCCs can merge at time $t+1$. Moreover, if $i < j$ and if there does not exist a directed path from $H_j(t)$ to $H_i(t)$ in $C(G^+(X(t)))$ with length 1 or 2, then we have $S_{ji}(t) = 0$ and $S_{jk}(t)S_{ki}(t) = 0$ for all $i < k < j$. After simple calculation, we can show $S_{ji}(t+1) = 0$, that is, no edge can appear from a node in the subset $H_j(t)$ to a node in the subset $H_i(t)$ at time $t+1$ in the digraph $G^+(X(t+1))$. Overall, the claims (i)-(ii) hold true.

Regarding facts (iii)-(v), based on facts (i)-(ii) and the analysis above, no two SCCs can merge, and the number $n(t)$ of SCCs associated to $G^+(X(t))$ is non-decreasing. Since the number of the group of individuals is finite and equivalently the number of all nodes of $G^+(X(t))$ is finite, the number of SCCs of the positive digraphs will remain constant after some finite time τ. Consequently, the set of SCCs of the positive digraphs also remains unchanged after τ. □

APPENDIX B
SUPPORTING LEMMAS AND PROPOSITION FOR THEOREM 5.1

Lemma B.1 (Products of primitive matrices with positive diagonal). *If $A_1, A_2, \dots, A_{n-1}$ are primitive $n \times n$ matrices with positive diagonal entries, then $A_1A_2 \cdots A_{n-1} > 0$.*

Proof: It suffices to show that any vector $x_0 \geq 0$, $x_0 \neq 0$, satisfies $A_{n-1} \cdots A_2A_1x_0 > 0$. For $t \in \{1, \dots, n-1\}$, let $x_t = A_tx_{t-1}$ and let n_{t-1} be the number of strictly positive entries in x_{t-1}. If $n_{t-1} < n$, then a suitable reordering of the entries leads to

$$x_{t-1} = \begin{bmatrix} p \\ 0 \end{bmatrix}, \text{ where } p \in \mathbb{R}^{n_{t-1}} \text{ and } p > 0.$$

Because A_t is primitive with positive diagonal entries, there exists $\epsilon > 0$ such that $A_t = \epsilon I_n + B$ for an appropriate primitive $n \times n$ matrix B. Hence,

$$x_t = A_t \begin{bmatrix} p \\ 0 \end{bmatrix} = \epsilon \begin{bmatrix} p \\ 0 \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} p \\ 0 \end{bmatrix}.$$

Note that (i) the first n_{t-1} entries of x_t are lower bounded by $\epsilon p > 0$, and (ii) the other entries of x_t are $B_{21} p \geq 0$. Because B is irreducible, B_{21} cannot be zero and, therefore, at least one entry of $B_{21} p$ is strictly positive. In summary, we have established that either $n_{t-1} = n$ or $n_t > n_{t-1}$. This proves the statement. □

Lemma B.2 (Convergence of appraisals in sinks)**.** *For the coevolution system* (4)-(5) *with $\epsilon > 0$, let $X(t)$ be an evolution with constant SCCs for time $t \geq \tau$. Let H_s be a sink of the condensation digraph with N_s nodes at time τ. If all positive appraisals held by individuals in H_s about individuals in H_s are lower-bounded by a positive constant β_s for all $t \geq \tau$, then the corresponding N_s rows of $X(t)$ converge exponentially fast to $\mathbb{1}_{N_\mathrm{s}} v^T$ for some $v \in \mathbb{R}^N$, where $v_j > 0$ if node j belongs to H_s and $v_j \leq 0$ otherwise.*

Proof: Denote the block influence matrix $W_\mathrm{s}(t)$ and appraisal matrix $X_\mathrm{s}(t)$ corresponding to a sink SCC H_s by $W_\mathrm{s}(t) \in \mathbb{R}_{\geq 0}^{N_\mathrm{s} \times N_\mathrm{s}}, X_\mathrm{s}(t) \in \mathbb{R}^{N_\mathrm{s} \times N_\mathrm{s}}$, $t \geq \tau$. Because all influences from the outside of the SCC are zero, we only need to study the coevolution within the SCC, i.e., $X_\mathrm{s}(t+1) = W_\mathrm{s}(t) X_\mathrm{s}(t)$, for all $t \geq \tau$. It is clear that $X_\mathrm{s}(t) = \prod_{k=1}^{t-\tau} W_\mathrm{s}(t-k) X_\mathrm{s}(\tau)$ for all $t > \tau$. Additionally, the maximum of each column of $X_\mathrm{s}(t)$ is always positive as $G^+(X_\mathrm{s}(t))$ is strongly connected, and the maximum is non-increasing for all $t \geq \tau$ due to the convex combination mechanism in the DeGroot model. Hence, from the definition (6), all positive entries of $W_\mathrm{s}(t)$ are lower-bounded for all $t \geq \tau$. Note that all matrices $W_\mathrm{s}(t)$, $t \geq \tau$, are row-stochastic, irreducible and, aperiodic; moreover, they have strictly positive diagonal entries and all their positive entries are lower bounded. Therefore, a direct application of Theorem 1 in [38] or, more specifically, Corollary 9.1 in [39] establishes the exponential convergence of $\prod_{k=1}^{t-\tau} W_\mathrm{s}(t-k)$ to a rank-1 matrix with identical rows.

Next, we claim that $X_\mathrm{s}(t) > 0$ for all sufficiently large t. We prove this claim by contradiction: For any sufficiently large t, there exists a $t' \geq t$ such that one entry $x_{ij}(t')$ of $X_\mathrm{s}(t')$ is equal to $\lambda \leq 0$. Then, based on the fact that $\prod_{k=1}^{t-\tau} W_\mathrm{s}(t-k)$ exponentially converges to a rank-1 matrix with identical rows, there exists a sufficiently small $\alpha > 0$ such that the values of all other entries in the j-th column of $X_\mathrm{s}(t')$ belong to $[\lambda - \alpha, \lambda + \alpha]$. For $\lambda < 0$, we can choose t so large that $\lambda + \alpha < 0$ and, hence, all entries in the j-th column of $X_\mathrm{s}(t')$ are negative. This fact implies that the j-th node is reducible for all subsequent times, which is a contradiction. For $\lambda = 0$, we can choose t so large that the entries of the j-th of $X_\mathrm{s}(t')$ are sufficiently close to 0 and such that $\alpha < \beta_\mathrm{s}$: if all entries are non-positive, then the j-th node is reducible which is a contradiction; if any one of them is positive, then it must be less than α and β_s, which contradicts the lower boundedness assumption on the positive entries of $X_\mathrm{s}(t)$. As a result, there exists a finite time t such that $X_\mathrm{s}(t) > 0$ and hence $W_\mathrm{s}(t) > 0$, which implies $X_\mathrm{s}(t+k) > 0$ for all $k \in \mathbb{N}$. Overall, the appraisal trajectory $X_\mathrm{s}(t)$ exponentially converges to $X_\mathrm{s}^* := \lim_{t \to \infty} X_\mathrm{s}(t)$, where $X_\mathrm{s}^* > 0$ is a rank-1 matrix with identical rows, by a direct application of Corollary 9.1 in [39]. The positive digraph associated with X_s^* is then fully connected.

For the remaining individuals in the group that do not belong to H_s at time τ, the influences from them to the sink are zero and the appraisals of them held by the sink nodes are non-positive. The N_s row of $\prod_{k=1}^{t-\tau} W(t-k)$, corresponding to the accumulated influences accorded by the sink H_s, then exponentially converges to $\mathbb{1}_{N_\mathrm{s}} w^T$ for some vector $w \in \mathbb{R}_{\geq 0}^{N_\mathrm{s}}$, where $w_j > 0$ if j is in the sink and $w_j = 0$ otherwise. Correspondingly, the associated N_s rows of $X(t)$ exponentially converges to $\mathbb{1}_{N_\mathrm{s}} v^T$ for some vector $v \in \mathbb{R}^{N_\mathrm{s}}$, where $v_j > 0$ if j is in the sink and $v_j \leq 0$ otherwise. Moreover, for a special case that $X(0) \geq 0$, it is straightforward to check $v \geq 0$. □

The lower-boundedness assumption in Lemma B.2 is a sufficient condition ensuring that the product of the influence submatrices $W_\mathrm{s}(t)$ (associated with the sink) converges to a rank-1 matrix. A weaker sufficient assumption is that all truncated positive appraisal submatrices (or equivalently, all truncated influence submatrices) associated with the sink are irreducible. Here, a *truncated matrix* is defined as follows: Given a constant $\beta \in \mathbb{R}_{>0}$, $x \in \mathbb{R}$, and $A = [a_{ij}] \in \mathbb{R}^{N \times M}$, we write $x^{\beta+} = \begin{cases} x, & \text{if } x \geq \beta, \\ 0, & \text{otherwise,} \end{cases}$ and $A^{\beta+} := [a_{ij}^{\beta+}]$. The matrix $A^{\beta+}$ is called the *truncation* of A by β. The digraph associated with $A^{\beta+}$ is called a *truncated digraph*. The sequence of signed matrices $\{A(t)\}_t$ has a *uniformly-irreducible truncation* if there exists a single $\beta > 0$ such that all matrices $A^{\beta+}(t)$ are irreducible.

Proposition B.3 (Convergence of appraisals in uniformly-irreducible truncated sinks)**.** *For the coevolution system* (4)-(5) *with $\epsilon > 0$, let $X(t)$ be an evolution with constant SCCs for time $t \geq \tau$. Let H_s be a sink of the condensation digraph with N_s nodes at time τ and with $N_\mathrm{s} \times N_\mathrm{s}$ appraisal submatrix $X_\mathrm{s}(t)$. If there exists a positive constant β_s such that $X_\mathrm{s}^{\beta+}(t)$ is irreducible for all $t \geq \tau$, then the corresponding N_s rows of $X(t)$ converge exponentially fast to $\mathbb{1}_{N_\mathrm{s}} v^T$ for some $v \in \mathbb{R}^N$, where $v_j > 0$ if node j belongs to H_s and $v_j \leq 0$ otherwise.*

Proof: Since $X_\mathrm{s}^{\beta+}(t)$ is irreducible for all $t \geq \tau$, there exists a constant $\gamma > 0$ such that $W_\mathrm{s}^{\gamma+}(t)$ is irreducible and with positive diagonal entries for all $t \geq \tau$. It is clear that $W_\mathrm{s}(t) \geq W_\mathrm{s}^{\gamma+}(t)$ for all t. Moreover, as $W_\mathrm{s}^{\gamma+}(t)$ is primitive and with positive diagonal entries, we have $\prod_{k=1}^{N_s-1} W_s^{\gamma+}(t+N_s-1-k)$ is positive for all $t \geq \tau$ by Lemma B.1. Hence, $\prod_{k=1}^{N_s-1} W_s(t+N_s-1-k)$ is positive and row-stochastic for all $t \geq \tau$, which implies the exponential convergence rate of the product of $W_s(t)$ to a rank-1 positive matrix by an application of Corollary 9.1 in [39]. The remaining proof is identical to that of Lemma B.2. □

As $X_s(t)$ of the sink SCC described in Proposition B.3 converges to a rank-1 positive matrix, the lower-bounded assumption on all positive entries of the appraisal matrices $X_s(t), t \geq \tau$, hold naturally.

Appendix C
Proof of Theorem 5.1

Consider the coevolution system (4) and (5) with $X(0) \in \mathbb{R}^{N\times N}$, a constant set of $SCCs$ beginning at time τ, and with $0 < \epsilon < 1$. It is useful to introduce the following definition. Given an SCC $H_k(t)$ with the corresponding appraisal square submatrix $X_k(t)$, consider a positive appraisal in $X_k(t)$, i.e., a positive entry $(X_k(t))_{ij}$. The *normalized positive appraisal* corresponding to $(X_k(t))_{ij}$ is $(X_k(t))_{ij}/\max_{h,\ell}(X_k(t))_{h\ell}$ and the *normalized positive appraisal submatrix* is $X_k^+(t)/\max_{h,\ell}(X_k(t))_{h\ell}$.

First, we make the following claims:

(A1): Consider a sink SCC H_s of $G^+(X(\tau))$. If the sequence of corresponding appraisal submatrices has a uniformly-irreducible truncation, then H_s converges exponentially fast to a faction. Instead, if there does not exist a single constant guaranteeing uniform irreducibility, then, for all $\beta > 0$, H_s includes multiple β-truncated (sink, intermediate or source) SCCs, each of which converges to factions, followers, or outsiders in $G^+(\lim_{t\to\infty} X(t))$.

(A2): Consider an intermediate (resp. source) SCC H_i of $G^+(X(\tau))$. If the positive appraisals of at least one sink SCC of $G^+(X(\tau))$ held by H_i are uniformly lower-bounded by a positive constant for all $t \geq \tau$, and if the sequence of normalized positive appraisal submatrices associated with H_i has a uniformly-irreducible truncation, then H_i converges asymptotically to the followers of the faction(s). Otherwise, the intermediate (resp. source) SCC may become factions or outsiders in $G^+(\lim_{t\to\infty} X(t))$.

Regarding (A1), the exponential convergence of a sink SCC H_s, whose truncation is uniformly irreducible, to a faction follows directly from Proposition B.3. If there does not exist a positive constant such that all truncated positive appraisal submatrices associated with the sink are irreducible, then the following two cases may occur.

(A1.a) *(truncated reducible case with globally reachable nodes)* Assume there exists $\beta > 0$ such that, for all $t \geq \tau$, the β-truncated appraisal digraphs of the sink H_s have a non-empty maximal set of globally reachable nodes (called *truncated globally reachable nodes*). Then, by the arguments similar to the ones in the proof of Proposition B.3, the rows of $X(t)$ corresponding to the appraisals held by the N_g truncated globally reachable nodes in H_s exponentially converges to $\mathbb{1}_{N_g} v^T$ for some vector $v \in \mathbb{R}^N$, where the j-th entry $v_j > 0$ if node j in H_s is truncated globally reachable and $v_j \leq 0$ otherwise. That is, the set of truncated globally reachable nodes converges to a faction. Note that the set of globally reachable nodes in the uniform truncation remain globally reachable in the limit and note that no other node in the sink H_s can become globally reachable in the limit. The convergence of the appraisals held by the other nodes of the sink H_s is established in the proof of (A2).

(A1.b) *(truncated reducible case with multiple sinks)* As the coevolution system (4) and (5) is continuous and has the convex combination feature of the DeGroot model, for any SCC H of $G^+(X(\tau))$ there exist a minimum positive constant β and a time $t_1 \geq \tau$ such that the β-truncated positive appraisal digraphs of H contain a constant set of SCCs for all $t \geq t_1$. We call these components of H the *truncated SCCs*. If the truncated SCCs of a sink H_s contains $K \geq 2$ sinks, then the rows of $X(t)$ corresponding to the appraisals held by the N_k nodes in the truncated sink SCC H_k, $k \in \{1,\dots,K\}$, converges to $\mathbb{1}_{N_k} {v^k}^T$, for some vector $v^k \in \mathbb{R}^N$. If H_k converges to a faction, then the j-th entry of v^k is greater than 0 for node j belonging to H_k and non-positive otherwise. If H_k includes only one node and her self-appraisal converges to 0 or is non-positive, then H_k converges to an outsider. The convergence of the truncated sinks to factions are similar to that of the truncated globally reachable nodes in the case (A1.a). (By comparison, (A1.a) is the case that $K = 1$.) The convergence analysis of the appraisals held by the remaining nodes in H_s is postponed to the proof of (A2).

Regarding (A2), for the simplicity of analysis, we relabel the nodes in such a way that the influence matrix $W(\tau)$ is block lower triangular as in equation (7) (recall that the influence matrix and the positive appraisal matrix have the same positive/zero pattern except the diagonal).

We start by considering an intermediate or source SCC H_i of $G^+(X(\tau))$ which is only directly connected to one sink SCC H_s for $t \geq \tau$. Assume H_s converges to a faction. We aim to show that H_i converges to singleton followers of the faction H_s subject to the assumptions in (A2).

Denote the block matrix corresponding to the influences within H_i as $D_i(t)$ and the block matrix corresponding to the influences of H_s to H_i as $S_{is}(t)$, (e.g., D_2 in equation (7) with $S_{21} \neq 0$). Note that $D_i(t)$ is row-substochastic for all finite $t \geq \tau$, and all positive entries of $S_{is}(t)$ can be proved to be lower-bounded by a constant $\gamma_1 > 0$ for all $t \geq \tau$ (by the lower-bounded positive appraisal assumption and by the arguments similar to the ones in the proof of Lemma B.2).

For $k \geq 0$, let $P_i(k,t) := \prod_{\ell=0}^{k} D_i(k-\ell+t)$. Since $D_i(t)$ is primitive (irreducible and aperiodic) with positive diagonal entries for $t \geq \tau$, Lemma B.1 implies that there exists a finite $d \geq 0$ such that any product $P_i(d,t) > 0$ for all $t \geq \tau$.

By assumption the normalized positive appraisals of $D_i(t)$ are lower-bounded by a positive constant for all $t \geq \tau$. Therefore, for each row of $D_i(t)$, each positive entry divided by the row sum is lower-bounded by a positive constant uniformly. Hence, each positive entry of $D_i(t)/r_{\max}(D_i(t))$ is lower-bounded by a positive constant uniformly, where we let $r_{\max}(A) > 0$ denote the maximum row sum of a non-negative matrix A. Similarly, we let $r_j(A) > 0$ denote the j-th row sum of a non-negative matrix A. Simple calculations show that

$$\frac{P_i(k,t)}{r_{\max}(P_i(k,t))} \geq \frac{\prod_{\ell=0}^{k} D_i(k-\ell+t)}{\prod_{\ell=0}^{k} r_{\max}(D_i(k-\ell+t))}.$$

In turn this implies that each entry of $P_i(d,t)/r_{\max}(P_i(d,t))$ is greater than a positive constant γ_2 uniformly (independent of t). Therefore, as $P_i(d,t)$ is row-substochastic, each row sum of $P_i(d+1,t) = P_i(d,t+1)D_i(t)$ satisfies:

$$\begin{aligned}
&r_j(P_i(d+1,t)) \\
&\quad = \sum_{n=1}^{N_i} p_{jn} r_n(D_i(t)) \\
&\quad \leq \sum_{n=1}^{N_i} \frac{p_{jn}}{r_{\max}(P_i(d,t))} r_n(D_i(t)) \\
&\quad \leq \sum_{n=1, n\neq k}^{N_i} \frac{p_{jn}}{r_{\max}(P_i(d,t))} + \frac{p_{jk}}{r_{\max}(P_i(d,t))} r_k(D_i(t)) \\
&\quad \leq 1 - \frac{p_{jk}}{r_{\max}(P_i(d,t))} + \frac{p_{jk}}{r_{\max}(P_i(d,t))} r_k(D_i(t)) \\
&\quad \leq 1 - \gamma_1\gamma_2, \qquad \text{for all } j \in \{1,\dots,N_i\},
\end{aligned}$$

where p_{jk} denotes the (j,k)-th entry of $P_i(d,t)$. Here we used the fact that there always exists one row of $D_i(t)$ (say, the k-th row) such that the row sum $r_k(D_i(t)) \leq 1 - \gamma_1$ due to the lower boundedness of the positive entries of $S_{is}(t)$ and the row-stochasticity of the matrix $[S_{is}(t) \quad D_i(t)]$. In other words, the maximum row sum of $P_i(d+1,t)$ is strictly less than $1 - \gamma_1\gamma_2$ uniformly (independent of t), and $\lim_{k\to\infty} P_i(k,\tau) = \lim_{k\to\infty} D_i(k+\tau)\cdots D_i(\tau) = 0$ exponentially. As H_{s} converges to a faction, there exists a large time t_1 such that all appraisals among H_{s} at time t for all $t \geq t_1$ are positive and lower-bounded, and each row sum of $S_{is}(t)$ is positive and lower-bounded. Consequently, there exists a time $t_2 \geq t_1$ such that all appraisals accorded to H_{s} by H_i are positive for all $t \geq t_2$. Moreover, one can verify that the i-th diagonal block matrix of the aggregate influence $\prod_{\ell=0}^{k} W(k-\ell+t)$ is equal to $P_i(k,t)$ and the corresponding block appraisal matrix is non-positive when $t \to \infty$. This implies that, when $t \to \infty$, the SCC H_i splits into N_i singleton SCCs. If each singleton SCC in the limit is directly connected to only one sink SCC, then it is straightforward to check the equilibrium appraisals accorded by the singleton to all members of the group in $\lim_{t\to\infty} X(t)$ is identical to the appraisals accorded by any individual in the sink SCC H_{s} to the group. The appraisals held by the singletons at $t = \tau$ are inessential.

In the arguments above, we assumed that H_{s} is a sink SCC and that it converges to a faction when $t \to \infty$. One can extend the convergence results of H_i to the case in which H_{s} is a truncated sink SCC converging to a faction when $t \to \infty$. The analysis is similar and will not be repeated here. We only remark that the lower-boundedness assumption on positive appraisals of the (truncated) sinks held by H_i is necessary for the convergence.

Next, we analyze the appraisal evolution of the intermediate or source SCCs which are not directly connected to a (truncated) sink SCC at $t = \tau$. We aim to show that such an SCC converges to faction, followers or outsiders.

Let H_i denote one of these intermediate or source SCCs. We assume that there exists a $\beta > 0$ such that all β-truncated normalized positive appraisal matrices associated with H_i are irreducible for all $t \geq \tau$. As we discuss in Section 4, the topology of $G^+(X(t))$ may not be stable after $t \geq \tau$ and the corresponding influence links between SCCs may change. We consider the following circumstances.

(A2.a) If H_i is only directly connected to a (truncated) sink SCC H_{s} (which converges to a faction) after a finite time $t_1 > \tau$, then the analysis for the SCCs that are directly connected to a (truncated) sink after time τ can be applied to H_i. That is to say, if the positive appraisals of H_{s} held by H_i are lower-bounded away from 0 uniformly for all $t \geq t_1$, then the SCC H_i splits into followers when $t \to \infty$. The appraisals held by the followers are identical to those held by the faction. On the other hand, if the positive appraisals of H_{s} held by H_i converge to 0, then H_i is disconnected from H_{s} in the limit: H_i converges to either a faction or an outsider. The convergence analysis of the appraisals held by H_i is then similar to the convergence of a sink SCC to a faction as discussed in (A1) except the possibility that H_i converges to a singleton with a non-positive self-appraisal, i.e., an outsider. Note that if one positive appraisal of H_{s} held by H_i is lower-bounded away from 0 uniformly for all $t \geq t_1$, then all positive appraisals of H_{s} held by H_i shall be lower-bounded away from 0 uniformly, subject to the irreducible assumption on both the truncated normalized positive appraisal matrices associated with H_i and the truncated positive appraisal matrices associated with the (truncated) sink SCC H_{s}. Therefore, if the positive appraisals of H_{s} held by H_i are not lower-bounded away from 0 uniformly, they must converge to 0 by the continuity of the coevolution system.

(A2.b) If H_i is only directly connected to an intermediate SCC H_j and H_j is directly connected to another (intermediate or sink) SCC H_k after a finite time t, then two exclusive cases may occur. (1) If the accumulated influences that H_i accord to H_j go to infinity, then H_j holds positive-complete appraisals of H_i after a certain time and those appraisals are lower-bounded away from 0 uniformly. Consequently, the appraisals of H_k held by H_i shall be positive after a finite time. That is to say, H_j is directly connected to H_k after a finite time, which contradicts our hypothesis. (2) If the accumulated influences that H_i accord to H_j are finite, then the influences that H_i accord to H_j converge to 0. That implies that H_i is disconnected from H_j in the limit: H_i converges to either a faction or an outsider. The convergence analysis of the appraisals held by H_i is then similar to the convergence of a sink SCC to a faction as discussed in (A1) with the exception the possibility for H_i to converge to a singleton with a non-positive self-appraisal.

(A2.c) We can extend the analysis to any intermediate or source SCC H_i at time τ. Two cases may occur. (1) H_i is only directly connected to one (sink or intermediate) SCC after a finite time t. This case has been discussed in (A2.a) and (A2.b). (2) H_i is directly connected to multiple (sink or intermediate or both) SCCs for infinitely many times. In this case, H_i will either converge to a faction (if all positive appraisals of the directly connected SCCs held by H_i converge to 0 and if H_i is not a singleton with non-positive self-appraisals), or converges to an outsider (if all positive appraisals of the directly connected SCCs held by H_i converge to 0 and if H_i is a singleton with non-positive self-appraisals), or split into followers (if the positive appraisals of at least one directly connected SCC held by H_i are lower-bounded away from 0 uniformly and this directly connected SCC converges to a faction). If one directly connected SCC H_j of H_i converges to followers or outsiders, then the positive appraisals of H_j held by H_i converge to 0, which implies that H_j and H_i are not directly connected in $G^+(\lim_{t\to\infty} X(t))$. It is possible that H_i includes the followers of multiple factions in the limit. If a follower is connected to multiple factions, then, by the convex combination feature of the coevolution model (4)-(5), the appraisals held by the follower are the convex combinations of the appraisals held by these different factions.

Note that the arguments of (A2.c) can be extended to the case (A1.b). That is, a sink SCC, which is truncated reducible and with multiple truncated sinks as described in (A1.b), converges to a factions-followers-outsiders structure which may include multiple factions, singleton followers or outsiders. The analysis is similar and will not be repeated here. Moreover, the convergence results above can also be extended to the case (A1.a). In particular, as the sink SCC H_s described in (A1.a) has a set of truncated globally reachable nodes for all $t \geq \tau$, H_s remains connected in the limit. This implies that (1) the set of truncated globally reachable nodes in H_s become a single faction in the limit and (2) the other sink nodes of H_s, which are not globally reachable in the uniform truncation, become singleton followers of the faction in the limit. The appraisals held by the followers are then identical to those held by the faction.

Finally, for any intermediate or source SCC H_i, if there does not exist $\beta > 0$ such that all β-truncated normalized positive appraisal matrices associated with H_i are irreducible for all $t \geq \tau$, then the SCC will split into multiple SCCs in $G^+(\lim_{t\to\infty} X(t))$; this split is similar to the one we analyzed for a sink SCC in (A1.a)-(A1.b). Moreover, for each H_i, we can find a finite set of truncated SCCs, such that the truncated normalized positive appraisal matrices associated with them are irreducible for all $t \geq \tau$. Then we can treat each truncated SCC of H_i as an intermediate or source SCC we considered in (A2.a)-(A2.c) due to the continuity of the coevolution system (4)-(5). The truncated intermediate or source SCCs converge in a way that is similar to the convergence properties of intermediate or source SCCs in (A2.a)-(A2.c).

We conclude the proof by summarizing our findings. Recall from Theorem 4.1 that, given any initial appraisal matrix $X(0)$, there exists a time τ such that the SCCs of $G^+(X(t))$ remain unchanged for all $t \geq \tau$. From our previous analysis we know the following facts. (1) All invariant SCCs or their truncated SCCs converge asymptotically to either factions or followers or outsiders. (2) $X(t)$ converges asymptotically to an equilibrium X^*. (3) There may exist multiple disconnected components in $G^+(X^*)$, even if $C(G^+(X(\tau)))$ is weakly connected. (4) $G^+(X^*)$ may include sink SCCs (factions or outsiders) and singleton source SCCs (followers), but neither intermediate SCCs nor source SCCs with multiple nodes. (5) The appraisals of single individuals held by each individual of the faction reach consensus: positive appraisal consensus of the individuals in the same faction and non-positive appraisal consensus of the remaining individuals. (6) If a follower is only directly connected to one faction, then the appraisals held by this follower are identical to the consensus appraisals held by the faction. If a follower is directly connected to multiple factions, then the appraisals held by the follower are the convex combinations of the appraisals held by the different factions. (7) As an outsider is a singleton sink with non-positive self-appraisal, her appraisals of all individuals are non-positive. These statements complete the proof of the claims (i)-(iv) in Theorem 5.1. □

Appendix D
Supporting lemmas for Theorem 5.2

Lemma D.1 (Convergence for non-negative $X(0)$). *Consider the coevolution system* (4)-(5)*. If $X(0) \in \mathbb{R}_{\geq 0}^{N \times N}$ is irreducible, and $\epsilon > 0$, then with an exponential rate,*

$$\lim_{t\to\infty} X(t) = \mathbb{1}_N v^T, \quad \textit{for some vector } v \in \mathbb{R}^N_{>0}.$$

Proof: Given irreducible $X(0) \geq 0$ and $\epsilon > 0$, it is easy to check $X(t) \geq 0$ for all $t \geq 0$ and

$$\begin{aligned} &X(t+1) \\ &= \operatorname{diag}\big((X(t) + \epsilon I_N)\mathbb{1}_N\big)^{-1}\big(X(t)^2 + \epsilon X(t)\big). \end{aligned} \tag{12}$$

In order to show that the trajectory $X(t)$ of the coevolution system converges to a rank-1 of the form $\mathbb{1}_N v^T$, we first show that

(i) if $X(t)$ is irreducible, then $W(t)$ is primitive and $X(t+1)$ is irreducible.
(ii) $\min^+_{i,j} w_{ij}(t) \geq \delta > 0$ uniformly for all $t \geq 0$.

Here $\min^+_{i,j} w_{ij}$ refers to the minimum non-zero positive entry of a non-negative matrix W.

Regarding fact (i), if $X(t)$ is irreducible, then by definition $W(t) = W_\epsilon(X(t))$ is irreducible. Since $\epsilon > 0$, $W(t)$ is also aperiodic. Moreover, as $X(t+1)$ is generated by (12) with $\epsilon > 0$, the irreducible $X(t)^2$ and $X(t)$ lead to $X(t+1)$ irreducible. Consequently, for all $t \geq 0$, $W(t)$ is primitive and $X(t)$ is irreducible.

Regarding fact (ii), we claim that

ii.1) there exists a finite time t_1 such that $X(t_1) > 0$ and then $W(t_1) > 0$;
ii.2) for $k \geq t_1$, $\min_{i,j\in\{1,\dots,N\}} w_{ij}(k) \geq \delta_1$ with $\delta_1 > 0$.

By the definition (6), the diagonal entries of $W(t)$, $t \geq 0$, are always positive. Therefore, $\prod_{k=1}^{t_1} W(t_1 - k) > 0$ for $t_1 = N - 1$ by Lemma B.1. It follows that $X(t_1) = \prod_{k=1}^{t_1} W(t_1 - k)X(0) > 0$ as $X(0)$ is irreducible.

For the second claim (ii.2), we have $X(k) > 0$ and $W(k) > 0$, if $k \geq t_1$. As $W(k)$ is row-stochastic, we have $\min_{i,j\in\{1,\dots,N\}} x_{ij}(k+1) \geq \min_{i,j\in\{1,\dots,N\}} x_{ij}(k)$ and $\max_{i,j\in\{1,\dots,N\}} x_{ij}(k+1) \leq \max_{i,j\in\{1,\dots,N\}} x_{ij}(k)$ which implies for all $k \geq t_1$,

$$\min_{i,j\in\{1,\dots,N\}} w_{ij}(k) \geq \frac{\min_{i,j\in\{1,\dots,N\}} x_{ij}(t_1)}{\epsilon + N\max_{i,j\in\{1,\dots,N\}} x_{ij}(t_1)} := \delta_1.$$

Since t_1 is finite, $\delta := \min(\delta_1, \min^+_{i,j,t\leq t_1} w_{ij}(t)) > 0$ is well defined. Then $\min_{i,j} w_{ij}(k) \geq \delta > 0$ for all $k \geq t_1$ from the claim (ii.2) above. Hence, fact (ii) holds for all $t \geq 0$.

Overall, $X(t) = \prod_{k=1}^{t} W(t-k)X(0)$ for all $t \geq 1$. From facts (i) and (ii) above, and Corollary 9.1 in [39], we obtain $\lim_{t\to\infty}\prod_{k=1}^{t} W(t-k) = \mathbb{1}_N w^T$ exponentially for some $w \in \mathbb{R}^N_{>0}$ and therefore $\lim_{t\to\infty} X(t) = \mathbb{1}_N v^T$, exponentially with $v = X^T(0)w > 0$. □

Lemma D.2 (Convergence for $X(0)$ with a positive column). *Consider the coevolution system* (4)-(5)*. If $X(0) \in \mathbb{R}^{N\times N}$ has one column with all positive entries and $\epsilon \geq 0$, then with an exponential rate,*

$$\lim_{t\to\infty} X(t) = \mathbb{1}_N v^T, \quad \textit{for some vector } v \in \mathbb{R}^N.$$

Proof: Define $V_{\max - \min} : \mathbb{R}^N \to \mathbb{R}_{\geq 0}$ by

$$V_{\max - \min}(x) = \max(x_1, \dots, x_N) - \min(x_1, \dots, x_N)$$

Without loss of generality, we assume that the first column of $X(0)$ is strictly positive, and therefore, the first column of $W(0)$ is strictly positive, which we denote as $W_{(:,1)}(0) > 0$. We claim that:

(i) for each $j \in \{1, \dots, N\}$, $V_{\max - \min}(X_{(:,j)}(t))$ is non-increasing w.r.t. $t \geq 0$, and it is strictly decreasing if greater than 0;
(ii) $\min_{t\geq 0, i\in\{1,\dots,N\}} w_{i1}(t) \geq \delta > 0$, i.e., all entries of all vectors $\{W_{(:,1)}(t), t \geq 0\}$ are uniformly greater than a positive value δ.

Regarding to the first claim (i), define $\delta(t) = \min_{i\in\{1,\dots,N\}} w_{i1}(t)$. Since $X_{(:,1)}(0) > 0$, and all $W(t)$ are row-stochastic for $t \geq 0$, all entries of $X_{(:,1)}(t+1)$ are the convex combination of the entries of $X_{(:,1)}(t)$. Hence, $X_{(:,1)}(t) > 0$ and $W_{(:,1)}(t) > 0$ for all $t \geq 0$, which implies $\delta(t) > 0$ for all t.

For all $i, j \in \{1, \dots, N\}$,

$$\begin{aligned} x_{ij}(t+1) &= \sum_{k=1}^{N} w_{ik}(t)x_{kj}(t) \\ &= w_{i1}(t)x_{1j}(t) + \sum_{k=2}^{N} w_{ik}(t)x_{kj}(t) \\ &\leq w_{i1}(t)x_{1j}(t) + (1 - w_{i1}(t))\max(X_{(:,j)}(t)) \\ &\leq \delta(t)x_{1j}(t) + (1 - \delta(t))\max(X_{(:,j)}(t)), \end{aligned}$$

and similarly,

$$x_{ij}(t+1) \geq \delta(t)x_{1j}(t) + (1 - \delta(t))\min(X_{(:,j)}(t)).$$

That is to say,

$$\begin{aligned} &V_{\max - \min}\left(X_{(:,j)}(t+1)\right) \\ &\leq \delta(t)x_{1j}(t) + (1-\delta(t))\max\left(X_{(:,j)}(t)\right) \\ &\qquad -\delta(t)x_{1j}(t) - (1-\delta(t))\min\left(X_{(:,j)}(t)\right) \\ &= (1-\delta(t))V_{\max - \min}(X_{(:,j)}(t)). \end{aligned} \tag{13}$$

Therefore, if $V_{\max - \min}(X_{(:,j)}(t)) > 0$, then $V_{\max - \min}(X_{(:,j)}(t+1)) < V_{\max - \min}(X_{(:,j)}(t))$ for all $t \geq 0$, $j \in \{1, \dots, N\}$; if $V_{\max - \min}(X_{(:,j)}(t)) = 0$, then $V_{\max - \min}(X_{(:,j)}(t+k)) = 0$ for all $k \geq 1$.

Regarding the second claim (ii), define $p_j = \max_{i\in\{1,\dots,N\}}(x_{ij}(0), 0)$ for all $j \in \{1, \dots, N\}$. By the arguments of the first claim, $p_j \geq \max_{i\in\{1,\dots,N\}} x_{ij}(t)$ for all t. From the definition, we have

$$\delta(t) = \min_{i\in\{1,\dots,N\}} w_{i,1}(t) \geq \min_{i\in\{1,\dots,N\}} x_{i1}(0)/(\sum_{j=1}^{N} p_j + \epsilon).$$

Then $\delta = \min_{i\in\{1,\dots,N\}} x_{i1}(0)/(\sum_{j=1}^{N} p_j + \epsilon)$ is well defined and is strictly greater than 0 since $p_1 > 0$ and

$\min_{i\in\{1,\ldots,N\}} x_{i1}(0) > 0$. As a result, $\delta(t) \geq \delta$ for all $t \geq 0$.

Overall, it is straightforward to check that the statement (13) still holds for the constant δ, which implies the exponential convergence of the trajectory of the function $V_{\max-\min}(X_{(:,j)}(t))$ to 0, and implies the exponential convergence of $X(t)$ to a rank-1 matrix $\mathbb{1}_N v^T$. □

Appendix E
Proof of Theorem 5.2

The proof of the claim (i) is a direct application of Lemma D.2 and that $G^+(X_\mathrm{s})$ is irreducible for all $t \geq \tau$ guarantees that all entries of $X_\mathrm{s}(t)$ are positive in the limit. That is, $\lim_{t\to\infty} X_\mathrm{s}(t) = \mathbb{1}_{N_\mathrm{s}} v^T$ for $v \in \mathbb{R}^{N_\mathrm{s}}_{>0}$. The claim (ii) is a direct application of Lemma D.2 where the nodes with positive appraisals held by all are globally reachable. The claim (iii) is a direct application of Lemma D.1. And the claim (iv) is straightforward. □

Appendix F
Proof of Lemma 6.2

First, for $X(t) \sim [+]$ for any $t \geq 0$. It is clear by the definition of the influence matrix (5) that $W(X(t)) \sim [+]$. Consequently, $X(t+1) = W(X(t))X(t) \sim [+]$. That is to say, the structure $[+]$ is invariant under the coevolution dynamics (4)-(5).

Second, if $X(t) = \begin{bmatrix} D_1 & S_1 \\ S_2 & D_2 \end{bmatrix}$ with $D_1 \in \mathbb{R}^{N_1\times N_1}$, $D_2 \in \mathbb{R}^{(N-N_1)\times(N-N_1)}$, $S_1 \sim [-]$, and $S_2 \sim [-]$, then $W(t) := \begin{bmatrix} D_1' & 0 \\ 0 & D_2' \end{bmatrix}$, where $D_1' = (\epsilon I + (1-\epsilon)D_1^+)\,\mathrm{diag}(D_1^+\mathbb{1}_{N_1})^{-1} > 0$ and $D_2' = (\epsilon I + (1-\epsilon)D_2^+)\,\mathrm{diag}(D_2^+\mathbb{1}_{N-N_1})^{-1} > 0$. It is noted that we assume each row sum of D_1^+ or D_2^+ is strictly positive for the simplicity of presentation. It is always true for $N_1 > 1$ and $N - N_1 > 1$, but may not hold for $N_1 = 1$ or $N - N_1 = 1$. If it is not true and $D_1 \leq 0$, based on the definition of $\hat{W}(X)$, then we simply set $D_1^+ = 1$ (similarly for D_2 and D_2^+). Consequently,

$$X(t+1) = W(t)X(t) = \begin{bmatrix} D_1'D_1 & D_1'S_1 \\ D_2'S_2 & D_2'D_2 \end{bmatrix} \sim \begin{bmatrix} D_1 & - \\ - & D_2 \end{bmatrix}.$$

That is to say, the structure $\begin{bmatrix} D_1 & - \\ - & D_2 \end{bmatrix}$ is invariant under the coevolution dynamics (4)-(5). □

Appendix G
Proof of Lemma 6.3

We can come to the conclusion in a similar way as we did for the second classical balanced structure in Lemma 6.2. □

Appendix H
Proof of Lemma 6.4

Under the condition that all entries in the same column of the appraisal matrix have the same sign, it is clear that, by relabelling the nodes, X has the form as in (8). If $X(t) = X$, we have

$$W(t) \sim \begin{bmatrix} + & \ldots & + & 0 & \ldots & 0 \\ + & \ldots & + & 0 & \ldots & 0 \\ + & \ldots & + & 0 & \ldots & 0 \\ + & \ldots & + & + & \ldots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ + & \ldots & + & 0 & \ldots & + \end{bmatrix} \quad \text{and hence,}$$

$$X(t+1) \sim \begin{bmatrix} + & \ldots & + & - & \ldots & - \\ + & \ldots & + & - & \ldots & - \\ + & \ldots & + & - & \ldots & - \\ + & \ldots & + & - & \ldots & - \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ + & \ldots & + & - & \ldots & - \end{bmatrix} \sim X(t).$$

Therefore the sign structure X in (8) is invariant under the coevolution (4)-(5). □

Appendix I
Proof of Corollary 6.1

The results are directly from the convergence analysis of Lemma B.2 and Theorem 5.1, and the macro-structure analysis of Lemmas 6.2–6.4. □