Noname manuscript No.
(will be inserted by the editor)

Who Should Be My Friends?

Social balance from the perspective of game theory

Wiebe van der Hoek · **Louwe B. Kuijer** ·
Yì N. Wáng

the date of receipt and acceptance should be inserted later

Abstract We define *balance games*, which describe the formation of friendships and enmity in social networks. We show that if the agents give high priority to future profits over short term gains, all Pareto optimal strategies will eventually result in a balanced network. If, on the other hand, agents prioritize short term gains over the long term, every Nash equilibrium eventually results in a network that is stable but that might not be balanced.

Keywords Structural balance theory · Game theory · Nash equilibrium · Pareto optimality

1 Introduction

A *social network* consists of a number of agents and positive or negative relations between them. The agents could be countries, individuals or groups. A positive relation represents a friendship or alliance, while a negative relation represents an enmity or rivalry. Structural balance theory describes such networks, and was introduced by Heider [15,16] and later generalized by Cartwright and Harary [11,12,3]. It argues that certain patterns are likely to occur while other patterns are unlikely; the likely patterns are referred to as *balanced* while the unlikely ones are *unbalanced*. There is also empirical support for the assertion that networks tend towards balance, see for example [27,30], though a fully balanced network is not always (nor easily) reached [20].

Balance theory describes a network as a whole; it is claimed (quite convincingly) that networks usually become more balanced over time, but relatively little attention is paid to the actions and motivations of individual agents on the way towards balance. Here, we take

This is an extended version of the paper [18] that was presented at LORI 2019. We thank the anonymous reviewers of LORI and of this special issue for useful comments and suggestions. Yì N. Wáng acknowledges funding support by the National Social Science Fund of China (18ZDA290, 20&ZD047).

Wiebe van der Hoek
Department of Computer Science, University of Liverpool, UK

Louwe B. Kuijer
Department of Computer Science, University of Liverpool, UK

Yì N. Wáng
Department of Philosophy (Zhuhai), Sun Yat-sen University, P.R. China
Corresponding author, E-mail: ynw@xixilogic.org

a game-theoretic approach: we explicitly treat the tendency towards balance as evidence for a preference by agents for balanced states over unbalanced ones. This allows us to take a detailed look at how this tendency follows from the rational choices by the individual agents.

We introduce a class of *balance games*, which are multi-stage games where in each stage one agent updates its relationship with someone else, and all agents prefer being involved in balanced relations over unbalanced ones. We show that if the agents are sufficiently patient or far-sighted (i.e., if they evaluate future income with a sufficiently small discount), any Pareto optimal strategy profile will, with probability 1, eventually result in a balanced network. If the agents are less patient, the end result may not be a balanced network. We show that for sufficiently impatient or short-sighted agents, any subgame perfect Nash equilibrium strategy profile will, with probability 1, result in a network that need not be balanced but that is *stable*. The concept of *stability* was defined in [17, 19] and is related to but strictly weaker than the concept of balance.

The structure of the paper is as follows. We first give definitions for balance, stability and the balance game in Section 2, where we also present a few useful lemmas, give an example, and discuss related work. Then, in Section 3 we consider the case of patient agents, and show that for them every Pareto optimal strategy profile results in balance. In Section 4 we study the cases of impatient agents. We generalize these results to directed graphs that are complete in Section 5, and that can be incomplete in Section 6 where we also introduce a structural theorem that generalizes [13, Theorem 13.2]. In Section 7 we discuss some generalizations as well as some limitations of our results. We conclude in Section 8.

2 Definitions and Preliminaries

In this section we first provide definitions of social balance theory, including structural balance and stability. Many of these are from the literature (mainly [3, 17, 19]). We give examples and introduce some results which will be used in later proofs. We then move on to define a class of balance games and some relevant notions. We use an example to explain the idea of balance games. We then discuss related approaches.

In this paper, we shall study balance games based on different versions of definitions of a *social network* (*network* for short). While in all versions a network is regarded as an irreflexive graph, they may differ in being undirected or directed, complete or incomplete:

- In Sections 2–4 we define balance games on complete, undirected networks.
- In Section 5 we generalize the games to work on complete directed networks.
- Finally, in Section 6 we generalize the games to work on any directed networks.

We will not consider incomplete undirected networks, for reasons that we briefly discuss in Section 6.3.

2.1 Structural balance and stability

A *complete undirected network* (abbreviated as a *network* in Sections 2–4) is a pair (A, E) such that A is a finite set of agents (represented by vertices of a graph), and $E : \{\{i, j\} \subseteq A \mid i \neq j\} \to \{+, -\}$ is an edge function that assigns to each unordered pair of different agents a positive $(+)$ or a negative $(-)$ edge. For simplicity, for pairs of agents we write ij, ik, etc, and for triads we write ijk, ijl, etc. We only consider networks with at least three agents.

2.1.1 Balance

Given a network $N = (A, E)$, a triad ijk of N is called *balanced*, if the labels of its edges are of one of the types $+++$ or $+--$ up to isomorphism. So in a balanced triad there is an even number of negative edges. The *unbalanced* triads therefore have either of the other two types: $++-$ or $---$. A network is *balanced*, if all of its triads are balanced, and *unbalanced* otherwise.

In a triad of the type $---$, all three agents are enemies of one another. In that situation, it is likely that two of them will set aside their differences and unite against their common foe. Doing so would turn the triad into $+--$, which is balanced. In a triad $++-$, there is one agent i that is friends with both j and k, while j and k are enemies. It is then likely that one of two things will happen: either the mutual friendship with i will form a basis for reconciliation between j and k, resulting in the balanced triad $+++$, or the tension between j and k will force i to end its friendship with one of them, resulting in the balanced triad $+--$. So both types of unbalanced triad have a tendency to evolve into a balanced triad.

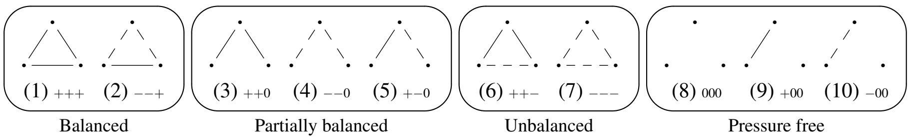


Fig. 1: The ten different triad shapes (up to isomorphism).

2.1.2 Stability

In addition to balance, we will also use the weaker notion of stability, which is defined in terms of mutual and anti-mutual ties. For a pair ij of a network $N = (A, E)$, a *mutual tie* of ij is an agent k of N such that k is a mutual friend or mutual enemy of i and j, i.e., either $E(ik) = E(jk) = +$ or $E(ik) = E(jk) = -$.

An *anti-mutual tie* of ij is an agent k of N such that k is either a friend of i and an enemy of j, or an enemy of i and a friend of j, i.e., if one of the following is true:

- $E(ik) = +$ and $E(jk) = -$
- $E(ik) = -$ and $E(jk) = +$.

We say a pair ij is *stable*, if it is one of the following cases (and *unstable* otherwise):

- $E(ij) = +$ and ij has at least as many mutual ties as anti-mutual ties;
- $E(ij) = -$ and ij has at least as many anti-mutual ties as mutual ties.

Finally, a network is *stable*, if all of its pairs are stable.

A mutual tie is a reason to stay or become friends, while an anti-mutual tie is a reason to stay or become enemies. A network is therefore stable if every pair of friends has at least as many reasons to remain friends as to become enemies, and every pair of enemies has at least as many reasons to remain hostile as to become friends.

2.1.3 Balance vs. stability

If ijk is a balanced triad and $E(ij) = +$, then k is a mutual tie for ij. Specifically, if ijk is of type $+++$ then k is a mutal friend, and if ijk is of type $+--$ then k is a mutual foe. Likewise, if ijk is balanced and $E(ij) = -$, then k is an anti-mutual tie for ij. A balanced network is therefore a stable network with the additional property that for all pairs ij, if $E(ij) = +$ then ij has only mutual ties and if $E(ij) = -$ then ij has only anti-mutual ties.

Not all stable networks are balanced, however. Two typical examples of stable networks that are not balanced are illustrated in Figure 2. In Figure 2(1), one can verify that every pair

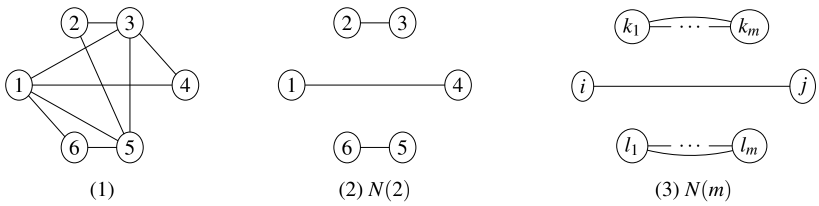


Fig. 2: Stable networks that are unbalanced, where a solid line stands for a positive edge and the lack of a line for a negative edge.

has an equal number of mutual and anti-mutual ties. For instance, pair $\{1,3\}$ has two mutual ties (i.e., agents 4 and 5) and two anti-mutual ties (i.e., agents 2 and 6). It is therefore stable, and so is the entire network. Yet the network is not balanced, for, e.g., the triad $\{1,2,3\}$ is not balanced. Similarly, the network of Figure 2(2) is also stable but not balanced.

The benefit of the latter network is that it can be generalized to a class of stable and unbalanced networks illustrated in Figure 2(3). For each natural number $m \geq 2$, the network $N(m)$ can be divided into three cliques: the $\{k_1,\ldots,k_m\}$-party (k-party for short), the $\{l_1,\ldots,l_m\}$-party (l-party for short) which are of equal size, and a small, third party $\{i, j\}$. Agents are friendly towards members of their own clique and hostile towards members of other cliques. The network shown in 2(2) is $N(2)$.

One can verify that for any pair $\{k_x,k_y\}$, $\{l_x,l_y\}$ or $\{i, j\}$ in the same party, there are $2m$ mutual ties (i.e., all others are their mutual ties), and is therefore stable. Any pair $\{k_x,l_x\}$ across the two major parties are stable, as there are 2 mutual ties (i.e., i and j) and $(2m-2)$ anti-mutual ties. Any pair $\{i,k_x\}$, $\{i,l_x\}$, $\{j,k_x\}$ or $\{j,l_x\}$ across the third party and a major party has an equal number (i.e., m) of mutual and anti-mutual ties, and is thus stable as well. For every $m \geq 2$, the network $N(m)$ is therefore stable. It is not balanced, however, because it contains triads of the type $---$.

Let us consider a few technical lemmas that will be useful later on. The first lemma is easy to obtain in balance theory, which follows immediately from the fact that a triad is balanced if and only if it contains an even number of negative edges.

Lemma 1 *If a triad ijk is balanced, then flipping (the sign of) any single edge of the triad will make it unbalanced. Likewise, if ijk is unbalanced then flipping any single edge of the triad will make it balanced.*

A pair ij is stable if and only if it is part of at least as many balanced triads as unbalanced triads. The following lemma therefore follows from Lemma 1.

Lemma 2 *If a pair ij is stable, then flipping $E(ij)$ does not increase the number of balanced triads containing i, nor does it decrease the number of unbalanced triads containing i.*

If a pair ij is unstable, then flipping $E(ij)$ will strictly increase the number of balanced triads in the network.

Finally, we need a *Relevance Lemma* that is less trivial and new in this paper.

Lemma 3 (relevance) *For any network, if there is an unbalanced triad, then all agents occur in an unbalanced triad.*

Proof If ijk contains an odd number of negative edges, then for every agent $l \notin \{i, j, k\}$ at least one of lij, ljk or lik also has an odd number of negative edges. □

The Relevance Lemma says that any unbalanced triad ijk is relevant to an agent l, since the existence of ijk implies that l itself is also involved in at least one unbalanced triad.

2.2 Balance games

We study structural balance from the viewpoint of game theory, by introducing a *balance game* which is a type of multi-stage game of infinitely many stages. All the agents in a network are players of a balance game. Each agent is better off if it is involved in more balanced triads. Players are allowed to update their ties with others, but only one player can update its relationship with one of the others in one step, and that takes one round or stage of a multi-stage game. Updates of relationships are made sequentially, which induces a sequence of networks, and this is common knowledge among all the players. Yet who will make a move in each step is completely non-deterministic.

Every stage game is based on a network, and different networks naturally yield different games. Moreover, we restrict to memoryless balance games, so that one network only yields one stage game (since players do not remember the previous moves, they have to make their moves based only on the network they are in). Therefore, in our setting there is a one-one correspondence between networks and stage games.

For a given network of players, the stage game for them is fixed. But since who will make a move is non-deterministic, there can be multiple succeeding stage games, each treated as being possible to happen under an equal probability. The entire game is composed of infinitely many stages, in a tree structure. Below we make this formal.

Valuation Given a network N, the valuation for an agent i in that network is the number of balanced triads i is part of minus the number of unbalanced triads it is part of. This valuation is denoted $val_i(N)$.

Actions At every stage, a single agent (chosen uniformly at random) will be given an opportunity to change one of its relations. This agent can choose to change its relation to one other agent, or it can choose to *pass* and leave all relations unchanged. Note that an agent can only change those relations that it is involved in. Agent i can decide to become enemies with j, but i cannot choose to create an enmity between j and k—although i might be able to create a situation where j and k have an incentive to become enemies. In a balanced network all triads are balanced, so balance is a *global* optimum of val_i for every i. In a stable network no single change to any relation ij would result in an increase in the number of balanced triads for either i or j (see Lemma 2), so stability is a *local* optimum of val_i for every i.

Discount factor At every stage of the game, the agents immediately receive utility equal to their valuation of the current network. This rewards them for having more balanced relations and punishes them for unbalanced ones. Additionally, they receive utility from future game stages. A reward today is worth more than the same reward tomorrow, however, so the agents multiply their future utility by a discount factor $\delta \in (0,1)$. The value of δ indicates the kind of agents that are being modeled; patient agents place (relatively) high value on the future and therefore have a high value for δ, impatient agents prioritize short term gain and therefore have a low value for δ. The utility for agent i in a network N therefore equals $val_i(N)$ plus δ times the expected utility in the successor network (if applicable, minus the *cost of change* that is explained below).

Cost of change If an agent decides to change a relation, it will incur a cost of change. This cost represents the effort and social cost associated with changing one's relation to another agent. For example, deciding to end an enmity might require an apology and a good bottle of wine, whereas ending a friendship may reduce one's social capital. The exact value that this cost of change can be debated. We believe that it should lie in the open interval $(0,2)$. In order to keep all calculations as simple as possible we prefer to have an integer cost of change, so we set it to be 1. See Section 7 for a discussion of why we believe that the cost of change should be between 0 and 2, and an overview of how any cost of change in the interval $[0,\infty)$ would influence our results.

We consider only memoryless pure strategies, so a strategy for an agent i can be represented by a function that maps every network to either a single change in a relation for i or to no change.[1] Below we introduce the formal definitions. We assume a fixed set of agents $A = \{1,\dots,n\}$ with $n \geq 3$, and use $\mathcal{N}$ to denote the set of all networks over A.

Definition 1 The *balance game* over a network $N = (A,E)$ is a pair (N,s) given by

- *(Players)* A is the set of players.
- *(Strategies)* $s = (s_1,\dots,s_n)$ is a strategy profile, such that for every player i, $s_i : \mathcal{N} \to \{(+,i,j),(-,i,j) \mid j \in A \setminus \{i\}\}$ is a strategy for i.
- *(Outcomes)* The outcome of (N,s) is one of $\{(N^{s_i},s) \mid i \in A\}$, chosen uniformly at random, where $N^{s_i} = (A,E^{s_i})$ is given by

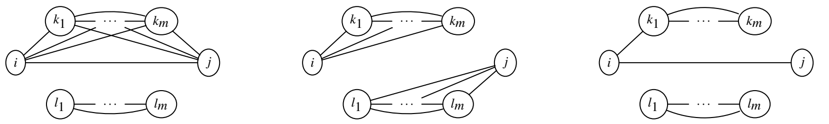
$$E^{s_i}(kl) = \begin{cases} +, & \text{if } s_i(N) = (+,i,j) \text{ and } kl = ij, \\ -, & \text{if } s_i(N) = (-,i,j) \text{ and } kl = ij, \\ E(kl), & \text{otherwise.} \end{cases}$$

- *(Utility)* The utility function $u = (u_1,\dots,u_n)$, where u_i is the utility of player i, is given recursively by $u_i(N,s) = val_i(N) + \delta \cdot \frac{1}{n} \cdot (\sum_{j \in A} u_i(N^{s_j},s) - c_j)$, where c_j – the *cost of change for* j – is such that $c_j = 1$ if $i = j$ and $N \neq N^{s_j}$, and $c_j = 0$ otherwise.

The recursive definition of utility does not immediately provide a practical way to compute $u_i(N,s)$. It is therefore useful to also have a direct characterization of $u_i(N,s)$. For this purpose, we use the concept of *timelines*. Given a strategy profile s, an *s-timeline* is an infinite sequence $l = \langle N_0,N_1,\dots\rangle$ such that for every $t \in \mathbb{N}$, $N_{t+1} \in \{N_t^{s_i} \mid i \in A\}$. The utility of agent i in such a timeline is given by $u_i(l) = \sum_{t=0}^{\infty} \delta^t (val_i(N_t) - c)$, where $c = 1$ if i brought about a change from N_{t-1} to N_t and $c = 0$ otherwise. The utility $u_i(N,s)$ is then simply the expected value of $\{u_i(l) \mid l = \langle N,N_1,\dots\rangle$ is an s-timeline$\}$.

For a given s-timeline $l = \langle N_0,N_1,\dots\rangle$, if there is a natural number T such that $N_{t_1} = N_{t_2}$ for all $t_1,t_2 \geq T$, then we say l *finalizes* in N_T, or N_T is the *final* of l.

[1] Neither of these restrictions is fundamentally necessary, all proofs presented in this paper can easily be adapted to mixed strategies that do use memory. But the restrictions do greatly simplify the proofs, so we assume them for reasons of clarity of presentation.



(1) A balanced outcome of $N(m)$ where i and j take the same side.

(2) A balanced outcome of $N(m)$ where i and j take different sides.

(3) A successor of $N(m)$ where i and k_1 become friends.

Fig. 3: Possible evolutions of the network $N(m)$ from Figure 2(3).

We write $N \rightsquigarrow_i N'$ if there is a strategy s_i for agent i such that $N' = N^{s_i}$, and we write $N \rightsquigarrow N'$ if there is at least one i such that $N \rightsquigarrow_i N'$.

As usual, we say a strategy profile is *Pareto optimal* (or simply, *optimal*) if there is no other strategy profile with which all players receive no less utility and at least one player gets a higher utility. A strategy profile is called a *subgame perfect Nash equilibrium* (or simply, an *equilibrium*), if no player could obtain a higher utility in any network by unilaterally changing its strategy.

2.3 Example

Consider the network $N(m)$ for a given $m \geq 2$ as depicted in Figure 2(3). In this network, most triads are balanced, but some remain unbalanced: the triads ikl and jkl are unbalanced for every $k \in \{k_1, \dots, k_m\}$ and every $l \in \{l_1, \dots, l_m\}$, since those triads are of the form $---$.

The agents could choose to pass, leaving the network in the state $N(m)$ forever. Alternatively, the agents can take actions that change the network. Taking such an action would incur a cost of change, however, so a rational agent will only do so in the expectation of a sufficiently high reward later. The main reward which all agents would like to obtain (though they may or may not be willing to pay the price for doing so) would be a balanced network.

There are many ways in which $N(m)$ can be changed to a balanced network. For example, all agents could decide to become friends with one another. That change would be very costly, however. Rational agents would instead aim for a balanced state that is easier to reach. A more feasible way to reach balance would be for the agents i and j to join the k-party or l-party, as shown in Figures 3(1) and 3(2).

Suppose that i joins the k-party. So eventually i will become friends with every agent k_x. Then at first, a friendship between i and some agent k_x must form. Without loss of generality, we can assume that this first friendship is with k_1, as shown in Figure 3(3). Consider the effect this has on the valuation of the different agents. Triads ik_1k_y and ik_1j used to be of the form $+--$ but are now $++-$. So they have turned from balanced to unbalanced. Triads ik_1l_z, on the other hand, used to be $---$ and have become $+--$, so they have turned from unbalanced to balanced. All other triads are unaffected. In total, there are $m-1$ triads ik_1k_y, 1 triad ik_1j and m triads ik_1l_z. So the number of triads that become balanced and the number of triads that become unbalanced are both m.

The agents i and k_1 are part of all triads that change, so their valuation is unchanged. One of them does have to pay the cost of change, but they suffer no harm from the change in the network. Agents l_y are part of one triad that changes, and it turns balanced. So their valuation increases, without them having to take any action. They quite like this change. The agents j and k_y are less happy, however: they too are part of one triad that changes, but theirs turns unbalanced. So they lose out due to this new friendship.

Once this first friendship has been established, all other members of the k-clique have an incentive to follow k_1 and become friends with i as well: currently, $k_1 k_y i$ is of the type $++-$, but by allying i they can turn this into the balanced type $+++$. So the first friendship ik_1 is likely to be followed by a flood of new friendships between i and the members of the k-party. Every such new friendship will be welcomed by the l-party, by i and by all k_y that are already friends with i, since it makes their relations more balanced. For those k_y that are not yet friends with i, the situation turns even worse, however. Every time an agent k_x becomes friends with i, the triad ik_yk_x becomes unbalanced, depriving k_y of another 2 points of valuation. In particular, if k_m is the last agent to become friends with i then just before they do so their valuation is $2(m-1)$ lower than it was in $N(m)$. Eventually, however, the network reaches one of the balanced states depicted in Figure 3, at which point all temporary losses are wiped away and replaced by the benefits of being part of a balanced network.

For highly impatient agents, paying the initial cost of change is not worth it, so remaining in $N(m)$ is the only rational option. If agents are more patient, however, aiming for balance may be the only rational choice. How patient agents have to be in order for remaining in $N(m)$ not to be an option depends on whether we are considering optimal strategy profiles or equilibria. The fact that the agents who are late to become friends with i (or j) suffer until balance is achieved means that remaining in $N(m)$ remains optimal until δ becomes very high. But the agents that experience a loss in valuation are not the ones that take action, it's the ones that have not yet taken action. So if the agents are even a little bit patient ($\delta = 0.5$ suffices, for example), the agents who decide to initiate the friendships will benefit by doing so, thereby making the strategy of remaining in $N(m)$ not an equilibrium.

2.4 Related work

Our definition of balance is called *3-balance* in the classical literature (e.g., [3]), where the number 3 refers to the length of the cycles to be examined – 3-cycles for triangles. In general, *k-balance* of a network requires that all cycles of length up to k contain an even number of negative edges. There is also pressure of balance from longer cycles, but it is considered of less effect [3]. This leads to a difference between viewing balance of networks as a *property* and as a *process*. Taking the former view, as in the classical literature, all cycles of all lengths are examined before we can decide the balance of the whole network. The lesser effect of longer cycles is modeled by assigning a *weight* or *strength* to each length [3,25]. In the latter view as proposed in [17] and adopted in this paper, however, the balance of a network lies in the balance of its local parts. The balance of longer cycles is achieved gradually over time by the constraints of balance among shortest cycles (triads in the case of undirected graphs).

The *structure theorem* [3,13] states that a balanced network can be partitioned into two mutually antagonistic and self-solidary components. The structure theorem was later generalized in [5] to consider a weaker version of balance which corresponds to more than two partitions. This gives a different way of studying the tendency of balance: it can be viewed as a process of partitioning a network. This approach has been developed in [7,8,26].

In recent years the study of link formation has drawn much attention in various fields including social network analysis, economics, information and computer science. Some of these are empirical studies that investigate, say, the formation of social networks or how technology is adopted in a network [31,4], and some are theoretical studies that focus on, say, the prediction, formal model, statistical and computational results of network formation [23,33,32,6,34,29]. This paper falls into theoretical side, and we focus on the formal model of a type of link formation from the viewpoint of game theory.

The study of structural balance theory has not been limited to a single field since the very beginning. It was initiated in Heider's work [15,16] in social psychology and reinvented by Harary et al. [11,12,3,13] using graph theory. Empirical studies on the impact of structural balance theory was carried out in the area of social network analysis (see, e.g., [27,28]). The trend to study and adopt the theory from new perspectives and in new fields has not come to an end. For example, the impact of structural balance on opinion formation has been evaluated in the framework of evolutionary games [22]. In our paper we also have structural balance and games in the same framework, but we focus more on the theoretical aspects of the structural balance of social networks.

Another area of related work is that of games on networks, a discipline of game theory concerned with networks. See for example [24,9,21]. Balance games can be considered part of this field, but they differ significantly from the games that have been studied before. Other disciplines of game theory, such as coalition formation and evolutionary games (see, e.g., [35]), are also related to balance games but very different from a technical point of view.

3 Patient Players

We show that for sufficiently patient players, a Pareto optimal strategy profile finalizes in a balanced network with probability 1. First, however, we consider two lemmas.

Lemma 4 *Let s be optimal and N a balanced network. Then $N^{s_i} = N$ for every agent i.*

Proof Taking any action other than passing incurs a cost of change, so in an optimal strategy an agent can only take such an action if they expect that doing so will eventually increase the valuation for at least one agent. In a balanced network every agent already has the highest possible valuation, so when playing an optimal strategy every agent passes.

Lemma 5 *Let s be a strategy profile, N_0 a network and L the set of s-timelines starting in N_0 that do not finalize in balance. If L occurs in the game (N_0, s) with probability greater than 0, then there is a $\delta_{high} < 1$ such that for all $\delta > \delta_{high}$, s is not Pareto optimal.*

Proof Suppose towards a contradiction that s is Pareto optimal and that L occurs with probability $p > 0$. Let N_{goal} be any balanced network, and let s' be the strategy where every agent, when given the opportunity, change their relations to match the ones in N_{goal}. We will show that, for sufficiently high δ, s' Pareto dominates s.

Every agent is part of $b := \frac{(n-1)\cdot(n-2)}{2}$ different triads. In a balanced network, all triads are balanced, so every agent has a valuation of b. In every non-balanced network, every agent has a valuation of at most $b-2$, since by Lemma 3 every agent is part of at least one unbalanced triad. Furthermore, by Lemma 4, every timeline that contains a balanced network must finalize in that network. So every network in every timeline $l \in L$ has a valuation of at most $b-2$, for every agent. This means that the expected valuation at any point in time is at most $p\cdot(b-2)+(1-p)\cdot b$. We therefore have $u_i(N_0,s) \leq \sum_{t=0}^{\infty} \delta^t (p\cdot(b-2)+(1-p)\cdot b) = \frac{p\cdot(b-2)+(1-p)\cdot b}{1-\delta}$.

Now, let N be any network and let k be the number of edges that differ between N and N_{goal}. We will compute a lower bound $f(k)$ on the expectation of $u_i(N,s')$. If $k=0$ then $f(k) = \sum_{t=0}^{\infty} \delta^t \cdot b = \frac{b}{1-\delta}$. If $k>0$, then there are two possibilities: either the agent that is chosen to act still has one or more edges left to change and does so, or it has no changes left to make and passes. The first possibility occurs with a probability of at least $\frac{1}{n}$ and the second with probability at most $\frac{n-1}{n}$. The valuation of N is at worst $-b$, so the expected utility in

N is at least $f(k) = -b + \delta(\frac{1}{n}(f(k-1) - c) + \frac{n-1}{n} f(k))$. Solving for $f(k)$ yields $f(k) = \frac{-b+\frac{\delta}{n}f(k-1)-\frac{\delta}{n}c}{1-\delta\frac{n-1}{n}}$. It follows that $f(k) = \sum_{i=1}^{k} \frac{(-b-\frac{\delta}{n}c)\cdot\left(\frac{\delta}{n}\right)^{i-1}}{\left(1-\delta\frac{n-1}{n}\right)^{i}} + \frac{\left(\frac{\delta}{n}\right)^{k} f(0)}{\left(1-\delta\frac{n-1}{n}\right)^{k}}$. As δ approaches 1, the latter expression approaches $(-bnk - ck) + f(0) = (-bnk - ck) + \frac{b}{1-\delta}$. So the strategy profile s' pays a constant price $(-bnk - ck)$, but in return it gains b times $\frac{1}{1-\delta}$, whereas s avoids the constant price but multiplies $\frac{1}{1-\delta}$ by the lower amount $p(b-2) + (1-p)b$. For sufficiently high δ, we therefore have $u_i(N_0, s) < u_i(N_0, s')$ for every agent i, contradicting the optimality of s. □

We get the following theorem from the above lemmas.

Theorem 1 *For a given number of players, there exists a discount factor δ_{high} such that for every $\delta > \delta_{high}$ and every Pareto optimal strategy profile s the following hold:*

1. *Every s-timeline that contains a balanced network finalizes in that network;*
2. *For every N, the game (N, s) reaches a balanced network with probability 1.*

Note that the bound δ_{high} depends on the number of agents. In fact, $\lim_{n\to\infty} \delta_{high} = 1$, so the required amount of patience approaches 1 as the number of agents increases.

This can, for example, be seen from the network $N(m)$ depicted in Figure 2(3). In order for $N(m)$ to become balanced, the central two agents i and j need to join either the clique $k_1, \ldots, k_m$ or the clique $l_1, \ldots, l_m$. While i is in the process of joining a clique, those members of the clique that are not yet friends with i experience a loss in valuation equal to twice the number of agents that are already friends with i. This loss is temporary, but both its magnitude and its duration increase with the number of agents. The amount of patience needed for any "go to balance" strategy to beat the "everyone passes in $N(m)$" strategy for every agent therefore increase with m.

4 Impatient Players

Here we show that if the discount factor δ is sufficiently close to 0, then every subgame perfect Nash equilibrium finalizes in a stable state with probability 1.

Unlike the case for patient agents, where the bound depends on the number of agents, our bound δ_{low} for impatient agents is constant. The proofs below can be used to determine an exact bound, which is the solution to a long equation that works out to something slightly greater than $\frac{1}{10}$. This bound is not tight, however, so its exact value does not seem very important. We therefore use the approximation $\delta_{low} = \frac{1}{10}$ instead.

Lemma 6 *Let N_0 be a network, and let m be the maximum increase of valuation brought about by any action of agent i, i.e., $m = \max\{val_i(N_1) - val_i(N_0) \mid N_0 \rightsquigarrow_i N_1\}$. Then for any strategy profile s, any s-timeline $\langle N_0, N_1, N_2, \ldots\rangle$ and any $t \in \mathbb{N}$ we have $val_i(N_t) \leq val_i(N_0) + (m + 2t)t$.*

Proof Consider the same action carried out in N_0 and N_k. This action will make some triads balanced, while making others unbalanced. Since N_0 and N_k differ in at most k edges, the number of triads made balanced when performing the action in N_k is at most k higher than in N_0, and the number of triads made unbalanced is at most k lower.

Turning a triad balanced increases valuation by 2, turning it unbalanced decreases it by 2. So in N_k the action yields at most $2k + 2k$ more valuation than in N_0, where it yields at most m. So the increase in valuation from N_k to N_{k+1} is at most $m + 4k$. It follows that $val_i(N_t) \leq val_i(N_0) + \sum_{k=0}^{t-1}(m + 4k) \leq val_i(N_0) + m \cdot t + \frac{4t}{2} \cdot t = val_i(N_0) + (m + 2t)t$. □

Lemma 6 places an upper bound on how quickly an agent's valuation can increase. Importantly, while the bound depends on the maximum possible gain m that the agent could make at time 0, it does not depend on the total number of agents in the network.

Lemma 7 *Let $\delta \leq \frac{1}{10}$ and s a Nash equilibrium. Then at every game (N,s), none of the agents take any action that changes the network unless that action increases their valuation.*

Proof Let k be the largest loss in valuation that any agent is willing to inflict upon themselves in any equilibrium, and suppose towards a contradiction that $k > 0$. Gains and losses in valuation come in multiples of 2, so $k \geq 2$.

Now, let (N,s) be a subgame where i makes a move that causes a loss of k in valuation and, as in the previous lemma, let $m = \max\{val_i(N') - val_i(N) \mid N \rightsquigarrow_i N'\}$. Consider also the alternative strategy s'_i where i always (i) makes the move with the highest possible immediate increase in valuation or (ii) passes if no increase in valuation is available, and let s' be the strategy profile that differs from s only in that i plays s'_i instead of s_i.

Let $\langle N, N_0, N_1, \ldots \rangle$ and $\langle N, N'_0, N'_1, \ldots \rangle$ be any s- and s'-timelines, respectively, with the property that in N it is i's action that is executed. Furthermore, let l be the maximal possible gain in valuation for i in N_0. Undoing the action that led to N_0 yields k in valuation, while doing any other action will yield at most $m+4$ valuation. So $l \leq \max\{k, m+4\}$. Then, by Lemma 6, we have $val_i(N_t) \leq val_i(N_0) + (l+2t)t = val_i(N) - k + (l+2t)t$.

In the sequence $N'_0 \rightsquigarrow N'_1 \cdots$ agent i may lose valuation. But this loss is bounded by k per time step: the only way for i to lose more than 2 in valuation in a single step is if an edge ia is changed, and in that case agent a shares the same loss, and by assumption no agent is willing to lose more than k valuation. So $val_i(N'_t) \geq val_i(N'_0) - k \cdot t = val_i(N) + m - k \cdot t$.

Finally, note that in the worst case the s'-timeline may require agent i to pay 1 utility as cost of change in each time step, whereas in the best case s never requires i to pay the cost of change after i's first action. We therefore have

$$\begin{aligned}
\tfrac{u_i(s') - u_i(s)}{\delta} &\geq m + k + \textstyle\sum_{t=1}^{\infty} \delta^t (-1 + val_i(N'_t) - val_i(N_t)) \\
&\geq m + k + \textstyle\sum_{t=1}^{\infty} \delta^t (-1 + (m - kt) - (-k + (l+2t)t)) \\
&= m + k + \textstyle\sum_{t=1}^{\infty} \delta^t (m + k - 1 - (k + l + 2t)t) \\
&> m + k - \textstyle\sum_{t=1}^{\infty} \delta^t (k + l + 2)t^2 \\
&\geq m + k - (m + 2k + 6) \textstyle\sum_{t=1}^{\infty} \delta^t t^2.
\end{aligned}$$

Because $\delta \leq \frac{1}{10}$ we have $\sum_{t=1}^{\infty} \delta^t t^2 < \frac{1}{6}$ (the property that we use is that $\sum_{t=1}^{\infty} (\frac{1}{10})^t t^2 = \frac{110}{729} < \frac{1}{6}$). Therefore, for any $m \geq 0$ and $k \geq 2$, $m + k \geq \frac{1}{6}(m + 2k + 6) > (m + 2k + 6) \sum_{t=1}^{\infty} \delta^t t^2$. It follows that $u_i(s') > u_i(s)$, so s is not an equilibrium.

We have arrived at a contradiction, so the assumption that $k > 0$ must have been false, which proves the lemma. □

Finally, if some agent has a valuation increasing move available, then such a move will be taken by at least one agent.

Lemma 8 *Let $\delta \leq \frac{1}{10}$ and s a Nash equilibrium. Then in every subgame (N,s), if any agent has an available action that will increase its valuation, then at least one agent takes an action that increases its valuation.*

Proof Any action that increases valuation increases it by at least two, so the increase in valuation outweighs the cost of change, resulting in a short term increase in utility. We omit the detailed calculations, but reasoning similar to that used in the proof of the previous lemma can be used to show that $\delta \leq \frac{1}{10}$ suffices to make this short term increase in utility outweigh any possible reward of not taking the valuation increasing action. □

Theorem 2 *Let $\delta_{low} = \frac{1}{10}$. Then for any discount factor $\delta \leq \delta_{low}$ and any subgame perfect Nash equilibrium s, the following hold:*

1. *Every s-timeline that contains a stable network finalizes in that stable network;*
2. *For every N, the subgame (N,s) reaches a stable network with probability 1.*

Proof Clause 1 follows from Lemma 7, and clause 2 from Lemmas 7 and 8. □

5 Extending to Complete Directed Networks

So far, we have treated social networks as undirected graphs. So if i is a friend of j then j is also assumed to be a friend of i. This is a simplifying assumption that is often justified; asymmetric relations are theoretically possible but almost never last. After all, if j hates i then it will be very hard for i to remain a friend of j.

However, because i's relation to j and j's relation to i may occasionally differ for a short while, it may at times be useful to have a more complex model where asymmetric relations are possible. In this section, we therefore introduce a variant of balance games on *digraphs*. Since we already introduced a variant of the game we can omit some of the details. We emphasize only those points where the definition and results of this section differ from the ones in Sections 2–4.

5.1 Stability over complete directed networks

In this section we generalize the results introduced in the previous sections to complete directed networks. That is to say, a network in this section is treated as a pair $N = (A,E)$ such that A is a finite set of agents, and $E : \{(i,j) \in A \times A \mid i \neq j\} \to \{+,-\}$ is an edge function that assigns to each *ordered* pair of different agents a positive $(+)$ or negative $(-)$ edge. We still write ij, ik, etc. for pairs of agents, and ijk, ijl, etc. for triads, though in the case of a digraph an edge between agents has a direction, so ij is different from ji.

While an edge in an undirected network can be thought of as a relation between two agents, an edge from i to j in a directed network is perhaps better understood as i's *attitude towards* j. The sign of the edge is $+$ if it is a positive attitude, and $-$ if negative.

Before introducing a formal definition of stability, we first explain the idea by introducing all the possible cases of shapes. Unlike in the case of undirected networks where we considered only the shapes of triads, the relationship between a pair is already relevant for stability in the case of directed networks. There are three possible cases of relationships between a pair of agents, namely $++$, $+-$ and $--$. Typically, two agents tend to have the same attitude towards each other. If not, there is usually a pressure or motivation for at least one of them to make a change. After all, it is hard to be friends with someone who considers you a foe. So among the three cases, $++$ and $--$ are balanced, while $+-$ is not.

For triads, there are however more cases to be considered than in the case of undirected networks. We list and categorize them into balanced and unbalanced ones in Figure 4.

Now we introduce the definition of *stability* of a digraph. We make use of the notions of *attraction* and *repulsion*, which extend the notions of *mutual* and *anti-mutual ties* seen in Section 2.1.2. The attraction between i and j is the number of balanced pairs and triads the edge ij is part of, while the repulsion is the number of unbalanced pairs and triads the edge is part of. Intuitively, the attraction is the number of reasons for two to become (or remain) friends, while the repulsion is the number of reasons to become (or remain) enemies.

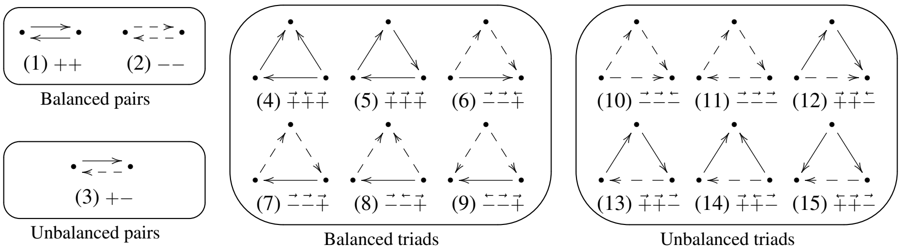


Fig. 4: The 15 different pair/triad shapes (up to isomorphism) for complete directed networks, where an arrow stands for a positive attitude, and a dasharrow for a negative attitude.

Definition 2 (stability) Let (A,E) be a social network, and let $i,k \in A$. The *attraction* of ik, denoted $attr(i,k)$, is given by $attr(i,k) = attr_2(i,k) + attr_3(i,k)$ such that:

$$\begin{aligned} attr_2(i,k) &= \begin{cases} 1, \text{ if } E(ki) = +, \\ 0, \text{ otherwise.} \end{cases} \\ attr_3(i,k) &= |\{j \mid E(ij) = E(jk) = +\}| + |\{j \mid E(ij) = E(jk) = -\}| \\ &+ |\{j \mid E(ij) = E(kj) = +\}| + |\{j \mid E(ij) = E(kj) = -\}| \\ &+ |\{j \mid E(ji) = E(jk) = +\}| + |\{j \mid E(ji) = E(jk) = -\}| \\ &+ |\{j \mid E(ji) = E(kj) = +\}| + |\{j \mid E(ji) = E(kj) = -\}|. \end{aligned}$$

The *repulsion* of ik, denoted $rep(i,k)$ is given by $rep(i,k) = rep_2(i,k) + rep_3(i,k)$ such that:

$$\begin{aligned} rep_2(i,k) &= \begin{cases} 1, \text{ if } E(ki) = -, \\ 0, \text{ otherwise.} \end{cases} \\ rep_3(i,k) &= |\{j \mid E(ij) = + \text{ and } E(jk) = -\}| + |\{j \mid E(ij) = - \text{ and } E(jk) = +\}| \\ &+ |\{j \mid E(ij) = + \text{ and } E(kj) = -\}| + |\{j \mid E(ij) = - \text{ and } E(kj) = +\}| \\ &+ |\{j \mid E(ji) = + \text{ and } E(jk) = -\}| + |\{j \mid E(ji) = - \text{ and } E(jk) = +\}| \\ &+ |\{j \mid E(ji) = + \text{ and } E(kj) = -\}| + |\{j \mid E(ji) = - \text{ and } E(kj) = +\}|. \end{aligned}$$

A pair ij is *stable* if it is one of the following cases (and *unstable* otherwise):

- $E(ij) = +$ and $attr(i,j) \geq rep(i,j)$;
- $E(ij) = -$ and $rep(i,j) \geq attr(i,j)$.

A network is *stable* if every pair of it is stable, and *unstable* otherwise.

The $attr_2$ and rep_2 components of attraction and repulsion represent the pressure on i's attitude towards j due to j's attitude towards i. The $attr_3$ and rep_3 components represent the pressure on i's attitude towards j due to the relations of i and j with third parties. It is easy to see from the definition that the $attr_3$ and rep_3 components are symmetric.

Proposition 1 *For any pair ij of a network, $attr_3(i,j) = attr_3(j,i)$ and $rep_3(i,j) = rep_3(j,i)$.*

5.2 Properties of stability over complete directed networks

Lemma 9 *All stable networks are symmetric networks (i.e., its edges are symmetric).*

Proof Given a stable network $N = (A, E)$ and a pair ij of N. Suppose $E(ij) = +$. Since N is stable, $attr(i, j) \geq rep(i, j)$. Suppose towards a contradiction that $E(ji) = -$. By definition $attr_2(i, j) = 0$ and $rep_2(i, j) = 1$. Thus, $attr(i, j) = attr_3(i, j)$ and $rep(i, j) = 1 + rep_3(i, j)$. On the other hand, $attr_2(j,i) = 1$ and $rep_2(j,i) = 0$, and so $attr(j,i) = 1 + attr_3(j,i)$ and $rep(j,i) = rep_3(j,i)$. By Proposition 1, $attr_3(i, j) = attr_3(j,i)$ and $rep_3(i, j) = rep_3(j,i)$. Therefore $attr(j,i) = attr(i, j) + 1$ and $rep(j,i) = rep(i, j) - 1$, and we have $attr(j,i) > rep(j,i)$. It contradicts with the stability of N. The case when $E(ij) = -$ is analogous. □

A symmetric network is close to an undirected network. It is not the case, however, that the attraction and repulsion of an edge in a symmetric directed network are exactly the same as the attraction and repulsion of the edge in the corresponding undirected network. This is because directed and undirected networks use slightly different methods for computing the attraction and repulsion of an edge. Fortunately, this difference turns out to be purely quantitative, as opposed to qualitative: an edge in a symmetric directed network is stable if and only if the corresponding edge in the undirected network is stable.

Definition 3 Given a symmetric directed network $N = (A, E)$, the *undirectification* of N is an undirected network $N' = (A, E')$ such that $E'(\{i, j\}) = E((i, j))$.

Theorem 3 *An edge (i, j) in a symmetric (complete, directed) network is stable if and only if the undirected edge $\{i, j\}$ is stable in its undirectification.*

Proof Given a symmetric network $N = (A, E)$, let N' be the undirectification of N. Suppose ij is a stable edge N and $E(ij) = +$, we have $attr(i, j) \geq rep(i, j)$. Observe that $attr(i, j) = 1 + 4 \cdot \# \{\text{mutual ties of } ij \text{ in } N'\}$ and $rep(i, j) = 4 \cdot \# \{\text{anti-mutual ties of } ij \text{ in } N'\}$. It follows that ij has no less mutual ties than anti-mutual ties, hence stable in N'. Similar arguments apply to the case where $E(ij) = -$. Suppose on the other hand that ij is a stable edge of N' and $E'(ij) = +$. The number of mutual ties of ij is not less than that of anti-mutual ties. By similar calculation we have $attr(i, j) > rep(i, j)$, and so ij is stable in N. We can show a similar result in the case when $E'(ij) = -$. □

Corollary 1 *A network is stable, iff it is symmetric and its undirectification is stable.*

5.3 Balance games over complete directed networks

Definition 4 The *balance game* over a complete directed network $N = (A, E)$ is a pair (N, s) such that:

- *(Players)* $A = \{1, \ldots, n\}$ (with $n \geq 2$) is the set of players.
- *(Strategies)* $s = (s_1, \ldots, s_n)$ is a strategy profile, such that for every player i, $s_i : \mathcal{N} \to \{(+, i, j), (-, i, j) \mid j \in A \setminus \{i\}\}$ is a strategy for i, where $\mathcal{N}$ is the set of all networks (digraphs) over A.
- *(Outcomes)* The outcome of (N, s) is one of $\{(N^{s_i}, s) \mid i \in A\}$, chosen at random, where $N^{s_i} = (A, E^{s_i})$ is given by
$$E^{s_i}(kl) = \begin{cases} +, & \text{if } s_i(N) = (+, i, j) \text{ and } (k, l) = (i, j), \\ -, & \text{if } s_i(N) = (-, i, j) \text{ and } (k, l) = (i, j), \\ E(kl), & \text{otherwise.} \end{cases}$$
- *(Utility)* Given a network N, player i's *valuation* in N, denoted $val_i(N)$, is the number of i's balanced pair or triad shapes in N minus the number of i's unbalanced pair or triad shapes in N. The *utility function* $u = (u_1, \ldots, u_n)$, where u_i is the utility of player i is given by $u_i(N, s) = val_i(N) + \delta \cdot \frac{1}{n} \cdot \sum_{j=1}^{n} (u_i(N^{s_j}, s) - c_j)$, where δ is a discount factor, $c_j = 1$ if $i = j$ and $N \neq N^{s_j}$, and $c_j = 0$ otherwise.

For explanations of the definition of balance games we refer to Section 2.2. The difference between a balance game over digraphs and that over undirected graphs are mainly in the values of the utility (counting pair and triad shapes instead of only undirected triads).

We find similar results of balance games, based on a distinction between patient and impatient players just as what was done in the case of undirected graphs.

Theorem 4 (decisions of patient players) *For a given number of players, there exists a discount factor δ_{high} such that the following hold for every $\delta > \delta_{high}$ and every Pareto optimal strategy profile s:*

1. *Every s-timeline that contains a balanced network finalizes in that network;*
2. *For every N, the game (N,s) reaches a balanced network with probability 1.*

Proof It can be shown using the same method as that for Theorem 1. □

Theorem 5 (decisions of impatient players) *Let $\delta_{low} = \frac{1}{34}$. Then for any discount factor $\delta \leq \delta_{low}$ and any subgame perfect Nash equilibrium s, the following hold:*

1. *Every s-timeline that contains a stable network finalizes in that stable network;*
2. *for every N, the subgame (N,s) reaches a stable network with probability 1.*

Proof Similarly to the proof of Theorem 2, let N_0 be a network and m be the maximum increase of valuation brought about by any action of agent i, then for any strategy profile s, any s-timeline $\langle N_0, N_1, N_2, \dots \rangle$ and any $t \in \mathbb{N}$, we have $val_i(N_t) \leq val_i(N_0) + (m+10t)t$ (a proof can be given similarly to that of Lemma 6). Using the same method as in the proof of Lemmas 7 and 8 (note that $m \geq 0$, $k \geq 2$ and $l \leq \max\{k, m+20\}$ in this case), we have that when $\delta \leq \frac{1}{34}$, in a Nash equilibrium strategy profile, no agent takes an action that does not increase its valuation, and at least one of them takes an action that increases it. □

6 Extending to Directed Networks That Can Be Incomplete

In this section, we generalize our results further to cover incomplete graphs. In this more complex variant, we allow a third type of edges representing the lack of a relationship, in addition to positive and negative ones. This new version of balance games is of the same style as the versions discussed previously, though significantly more complex, so we shall emphasize the differences from those based on complete graphs.

6.1 3-signed directed networks

We consider directed social networks that can be incomplete, which by definition are irreflexive *2-signed* digraphs (with "+" and "−" for the positive and negative signs, respectively). Yet, formally, we introduce a third sign "0" for the *lack of* an attitude[2], and treat a directed network as an irreflexive, *complete*, *3-signed* digraph.

Besides the balanced and unbalanced shapes studied previously, we see more cases in regard to the lack of attitudes. We illustrate the possible pair shapes and triad shapes in Figure 5. A pair in the shape $+-$ has a pressure to change to $++$ or $--$. Shape $+0$ has a

[2] The lack of an attitude may be due to an agent's ignorance or unawareness of the other. Occasionally we may also understand a 0-edge to be a neutral or indifference attitude.

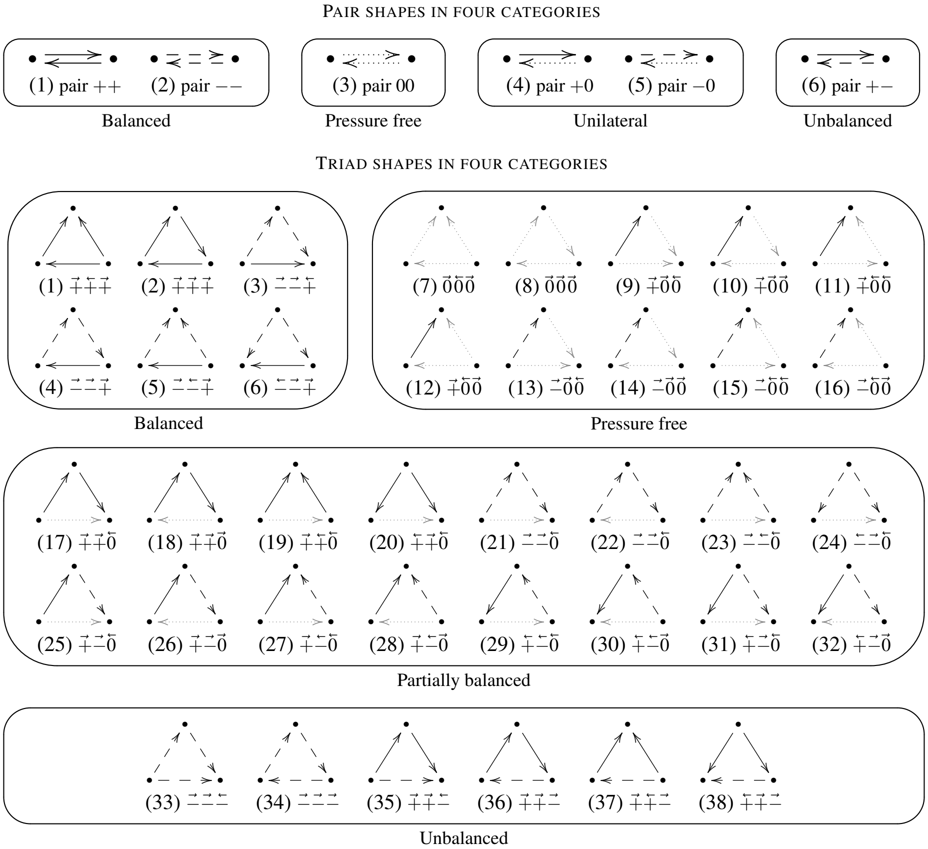


Fig. 5: The 6 different pair shapes and 38 different triad shapes (up to isomorphism) for directed networks that can be incomplete, where an arrow, a dasharrow and a dotted arrow stand for a positive, a negative and the lack of an attitude, respectively.

pressure to change to $++$[3], and -0 to $--$. For triads, we can likewise categorize them into four cases, the *balanced*, *unbalanced*, *partially balanced* and *pressure-free* ones. For pairs and triads, the balanced and pressure-free shapes (together, we call them *semi-balanced*) are not subject to change, while all or part of the agents involved in the other two types (we shall call them *semi-unbalanced*) has a reason to revise their attitudes. Moreover, we say a network is *semi-balanced* if all of its pair and triad shapes are semi-balanced.

Definition of *stability* of a 3-signed digraph needs to cover the cases of 0-edges. We can extend Definition 2 (for 2-signed digraphs) with an extra condition for the stability of a pair ij (and keeping other parts untouched):

– $E(ij) = 0$ and $attr(i, j) = rep(i, j)$.

We can also consider 3-signed undirected graphs, categorize its triad shapes into four cases similarly to that in Figure 5 (with only 10 different cases), and define stability for it like in Section 2.1.2 (a precise definition appeared in [17, 19]).

3 We consider it hard for someone to go from a positive or negative attitude towards someone to be ignorant of that person, so 00 is in general not a possible output of $+0$.

Theorem 6 (3-signed version of Theorem 3)

1. *An edge (i,j) in a symmetric 3-signed digraph is stable if and only if the undirected edge $\{i,j\}$ is stable in its undirectification.*
2. *As a corollary, a 3-signed digraph is stable, if and only if it is symmetric and its undirectification is stable (cf. [19, Definition 4]).*

6.2 Structural properties of stability

Now that we have introduced various concepts of networks/graphs, we would like to give a summary as follows:

balanced digraphs $\sqsubset$ semi-balanced digraphs $\sqsubset$ stable digraphs $\sqsubset$ symmetric digraphs

where $\sqsubset$ means that the concept on the left is strictly less general than the one on the right, and the above holds for 3-signed graphs (for 2-signed graphs, the above holds if we skip the concept "semi-balanced digraphs").

The above is not hard to see. By definition semi-balance is a more general concept than balance. Also, all stable digraphs are symmetric (Lemma 9 and Theorem 6), but not vice versa (an unstable symmetric digraph exists). Moreover, the following hold.

Proposition 2 *Every semi-balanced network is stable, but not necessarily vice versa.*

Proof To see that all semi-balanced networks are stable, all we need is to verify that all the shapes allowed are stable, and that is the case. For the converse direction, consider Figure 2(1), which is undirectedly stable (just treat every line as a pair of bidirectional arrows; recall that the lack of a line there stands for negative attitudes), and thus stable by Theorem 9, but yet it is not semi-balanced (check, say, the triad (1,4,5) which is unbalanced). □

In [13] it is shown (Theorem 13.2) that, for any network (i.e., 2-signed incomplete digraph) N, N is balanced[4] if and only if the vertices of N can be partitioned into two subsets (one of them may be empty) such that every positive edge joins two vertices of the same subset and every negative edge joins two vertices of different subsets. This is often referred to as *structure theorem* or *balance theorem*. We can show a parallel in terms of semi-balance.

Theorem 7 (structure theorem) *Given a network N, N is semi-balanced if and only if its edges are symmetric and the vertices of N can be partitioned into k ($k \geq 1$) subsets such that all of the following hold:*

1. *Every pair of vertices in the same subset are joined by a positive edge;*
2. *Either all edges between two subsets are negative, or all of them are neutral;*
3. *Every vertex cannot have negative edges to or from more than one different subsets.*

Proof Suppose N is semi-balanced. By definition it must be symmetric. Let $V_1, V_2, \ldots, V_k$ be the partition such that $a, b \in V_i$ (for some $i = 1, \ldots, k$) if and only if there is a path of positive edges from a to b via vertices in V_i. It is easy to observe from the definition of semi-balance that the positive edges are transitive, and so the first and second clauses hold. For the third clause, suppose there is a vertex that has two different negative edges to or from

[4] The notion of balance in [13] is defined for a less general concept, in the sense that our definition of balance in Section 5.1 (only for pairs and triads, but not longer cycles) is called *3-balance* there, and the balance defined there needs to be achieved for any length of cycles. For details see [13, p. 341].

two different subsets, it follows that the vertex is involved in one of the $---$ or $--0$ triad shapes (in any direction), which conflicts with the assumption that N is semi-balanced.

For the converse direction, suppose those conditions hold for a network N and we must show N is semi-balanced. By the symmetry of N we get that all pairs are of the shape $++$, $--$ or 00. Now for any triad abc, if a, b and c are in the same partition, then by the first clause the triad shape is $+++$ (in any direction). If two of them are in the same partition and the other in a different one, then by the conditions they are of the shape $+--$ or $+00$ (in any direction). If the three vertices are all in different partitions, then they are of the shape 000 or -00 (in any direction). In each case, abc is balanced or pressure free, hence semi-balanced.

6.3 Balance games over incomplete digraphs

As for the balance games over 3-signed digraphs, we can adopt Definition 4 by replacing the occurrences of "balanced" with "semi-balanced", and "unbalanced" with "semi-unbalanced". We can also further extend the definition to allow a player i performing an action $(0,i,j)$ such that i becomes ignorant of another player j (better understood as i changes to a neutral or indifferent attitude towards j in this case). We can show a 3-signed version of Theorems 4 and 5 using a similar proof method. Details are omitted.

We did not focus on 3-signed undirected graphs in this paper, due to a difficulty in defining an action $(0,i,j)$. To enforce symmetry, an agent is either allowed to break up without the feedback from others, or is forbidden to do so. While both are unnatural in reality, the former even leads to a fact that all players have an easy way to profit, namely to break up with all others. There does not seem to be an easy adaption to our framework that avoids this issue.

7 Discussion

Accuracy Balance theory predicts that social networks broadly tend towards balance, but that a fully balanced network is not always reached. This is also confirmed by empirical studies. The same general behavior is observed in balance games: rational agents will generally increase the amount of balance in the network, but under most circumstances a fully balanced outcome is not guaranteed.

Whether balance games accurately predict agents' behaviour on a more detailed level is not currently known, and remains an interesting question for further research.

Pareto optimality for low δ and subgame perfect Nash Equilibria for high δ Our results are "asymmetric", in the sense that δ_{high} is related to optimality while δ_{low} is related to equilibria. We conjecture that this asymmetry is fundamental: we think that for arbitrarily high $\delta < 1$ there remain equilibria that do not finalize in balanced networks and that for arbitrarily low $\delta > 0$ there remain Pareto optimal strategy profiles that do not finalize in stable networks. Unfortunately, the strategy space for balance games is very large and hard to describe. So while we have reasons to believe that there are no lower bound for optimality and upper bound for equilibria, we have not yet managed to find the counterexamples that prove this to be the case.

Cost of Change Changing a relation takes some amount of effort, so it should be associated with some cost $c > 0$. Furthermore, agents seem willing to incur this cost in order to make their relations more balanced. This suggests that the increase in valuation caused by the increase in balance is higher than the cost of change, so $c < 2$. We therefore consider values of c outside the interval $(0,2)$ to be implausible. Still, for the sake of completeness we explain how our results change for any $c \in [0,\infty)$.

The bound δ_{high} is not qualitatively affected by the cost of change: for every $c \in [0,\infty)$, there is still a bound δ_{high} above which every optimal solution finalizes in balance with probability 1 and δ_{high} approaches 1 as n approaches infinity.

For any $c \in (0,2)$, the bound δ_{low} is also qualitatively unaffected. The exact value of the bound may change, but a $\delta_{low} > 0$ still exists and is independent of the number of agents.

For $c \in (2,\infty)$, on the other hand, we do get different results. The first statement of Theorem 2 still applies: every equilibrium timeline that contains a stable network finalizes in that network. But the second part of Theorem 2 does not hold for $c \in (2,\infty)$. If $c > 2$ and δ is sufficiently low then some timelines finalize before reaching a stable network.

This leaves the two cases $c = 0$ and $c = 2$. If $c = 0$, then no bound δ_{low} exists: for every $\delta \in (0,1)$ there are equilibria where agents move out of a locally optimal stable state and eventually reach a globally optimal balanced state. Finally, for $c = 2$, there is a bound δ_{low}, but in that case we do not know whether $\lim\limits_{n\to\infty} \delta_{low} = 0$.

8 Conclusion

In this paper we viewed structural balance of a social network as a result of its agents playing a *balance game*. When the agents are patient, their Pareto optimal strategies result in a *balanced* network as the game proceeds. When the agents are impatient, their subgame perfect Nash equilibrium strategies result in a *stable* network. By a framework accommodating both the concepts of balance and stability, our work bridged the classical literature on social balance [3] and its recent development using a logical approach [17,19,34,29].

There is still work that remains to be done. In particular, while we have shown that bounds δ_{high} and δ_{low} exist, we have not yet found tight bounds. Furthermore, as mentioned in Section 7, we conjecture that an equilibrium for patient agents may not finalize in balance and that an optimal profile for impatient agents may not finalize in balance. A proof (or, for that matter, a disproof) of these conjectures would be interesting. It would also be good to know more about the behaviour of agents that are neither as patient as to guarantee balance nor so impatient to guarantee stability.

Additionally, there are a number of further questions related to generalizations of the balance game. The balance game could, e.g., be generalized to different kinds of networks, including weighted networks (where some friendships/enmities are stronger than others). It should also be interesting to allow different kinds of agents. Some agents might be more patient than others, or have a higher tolerance for unbalance. The framework of Boolean games [14,10] seems to be appropriate for modelling the diversity of agents in their goals.

Another way to increase diversity is in the strategies of agents. By going further to formalizing the dynamics of balance games in the framework of temporal logic, in particular, alternating-time temporal logic [1,2], we can get a better characterization of the time evolution and the flexibility of modeling agent's strategies in a formal and unified manner. We leave, however, all these for future work.

References

1. Alur, R., Henzinger, T.A., Kupferman, O.: Alternating-time temporal logic. In: Proc. FOCS '97, pp. 100–109. IEEE (1997)
2. Alur, R., Henzinger, T.A., Kupferman, O.: Alternating-time temporal logic. JACM **49**(5), 672–713 (2002)
3. Cartwright, D., Harary, F.: Structure balance: A generalization of Heider's theory. Psychological Review **63**(5), 277–293 (1956)
4. Conley, T.G., Udry, C.R.: Learning about a new technology: Pineapple in Ghana. American Economic Review **100**(1), 35–69 (2010)
5. Davis, J.A.: Clustering and structural balance in graphs. Human Relations **20**(2), 181–187 (1967)
6. Dev, P.: Networks of information exchange: Are link formation decisions strategic? Economics Letters **162**, 86–92 (2018)
7. Doreian, P., Mrvar, A.: A partitioning approach to structural balance. Social Networks **18**(2), 149–168 (1996)
8. Doreian, P., Mrvar, A.: Partitioning signed social networks. Social Networks **31**, 1–11 (2009)
9. Goyal, S.: Connections: An Introduction to the Economics of Networks. Princeton (2007)
10. Gutierrez, J., Harrenstein, P., Wooldridge, M.: Iterated boolean games. Information and Computation **242**, 53–79 (2015)
11. Harary, F.: On the notion of balance of a signed graph. Michigan Math. J. **2**(2), 143–146 (1953)
12. Harary, F.: On local balance and n-balance in signed graphs. Michigan Math. J. **3**(1), 37–41 (1955)
13. Harary, F., Norman, R.Z., Cartwright, D.: Structural Models: An Introduction to the Theory of Directed Graphs. John Wiley & Sons Inc, New York (1965)
14. Harrenstein, P., van der Hoek, W., Meyer, J.J., Witteveen, C.: Boolean games. In: Proc. of the 8th Conference on Theoretical Aspects of Rationality and Knowledge, pp. 287–298 (2001)
15. Heider, F.: Social perception and phenomenal causality. Psychological Review **51**(6), 358–374 (1944)
16. Heider, F.: Attitudes and cognitive organization. J. of Psychology **21**(1), 107–112 (1946)
17. van der Hoek, W., Kuijer, L.B., Wáng, Y.N.: A logic of allies and enemies. In: 13th Conference on Logic and the Foundations of Game and Decision Theory (LOFT). Italy (2018)
18. van der Hoek, W., Kuijer, L.B., Wáng, Y.N.: Who should be my friends? Social balance from the perspective of game theory. In: LORI 2019, LNCS 11813, pp. 370–384 (2019)
19. van der Hoek, W., Kuijer, L.B., Wáng, Y.N.: Logics of allies and enemies: A formal approach to the dynamics of social balance theory. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI) (2020)
20. Hummon, N.P., Doreian, P.: Some dynamics of social balance processes: bringing Heider back into balance theory. Social Networks **25**(1), 17–49 (2003)
21. Jackson, M.O., Zenou, Y.: Games on networks. In: Handbook of Game Theory, vol. 4, pp. 95–164 (2015)
22. Li, W., Li, P., Wang, H., Fan, P.: Evolutionary game of opinion dynamics under impact of structural balance. In: Proc. of the 3rd International Conference on System Science, Engineering Design and Manufacturing Informatization, pp. 208–211. IEEE (2012)
23. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. Journal of the American Society for Information Science and Technology **58**(7), 1019–1031 (2007)
24. de Martí, J., Zenou, Y.: Social networks. In: Handbook of the Philosophy of Social Sciences, pp. 339–361. SAGE Publications (2011)
25. Morrissette, J.O.: An experimental study of the theory of structural balance. Human Relations **11**(3), 239–254 (1958)
26. Mrvar, A., Doreian, P.: Partitioning signed two-mode networks. J. Math. Sociol. **33**, 196–221 (2009)
27. Newcomb, T.M.: Acquaintance Process. Holt, Rinehart & Winston, New York (1961)
28. Newcomb, T.M.: Reciprocity of interpersonal attraction: A nonconfirmation of a plausible hypothesis. Social Psychology Quarterly **42**(4), 299–306 (1979)
29. Pedersen, M., Smets, S., Ågotnes, T.: Analyzing echo chambers: A logic of strong and weak ties. In: Proceedings of LORI 2019, pp. 183–198 (2019)
30. Sampson, S.F.: A novitiate in a period of change: An experimental and case study of social relationships. Ph.D. thesis, Cornell University (1968)
31. Santos, P., Barrett, C.B.: Identity, interest and information search in a dynamic rural economy. World Development **38**(12), 1788–1796 (2010)
32. Smets, S., Velázquez-Quesada, F.R.: How to make friends: A logical approach to social group creation. In: Logic, Rationality, and Interaction, pp. 377–390. Springer (2017)
33. Stattner, E.: Involvement of node attributes in the link formation process into a telecommunication network. Social Network Analysis and Mining **5**(1), 64 (2015)
34. Xiong, Z., Ågotnes, T.: On the logic of balance in social networks. JoLLI **29**, 53–75 (2019)
35. Young, H.P., Zamir, S. (eds.): Handbook of Game Theory, Volume 4. North-Holland (2015)