Dynamical Models Explaining Social Balance and Evolution of Cooperation

Vincent Antonio Traag[1]*, Paul Van Dooren[1], Patrick De Leenheer[2]

[1] ICTEAM, Université catholiquede Louvain, Louvain-la-Neuve, Belgium, [2] Department of Mathematics, University of Florida, Gainesville, Florida, United States of America

Abstract

Social networks with positive and negative links often split into two antagonistic factions. Examples of such a split abound: revolutionaries versus an old regime, Republicans versus Democrats, Axis versus Allies during the second world war, or the Western versus the Eastern bloc during the Cold War. Although this structure, known as social balance, is well understood, it is not clear how such factions emerge. An earlier model could explain the formation of such factions if reputations were assumed to be symmetric. We show this is not the case for non-symmetric reputations, and propose an alternative model which (almost) always leads to social balance, thereby explaining the tendency of social networks to split into two factions. In addition, the alternative model may lead to cooperation when faced with defectors, contrary to the earlier model. The difference between the two models may be understood in terms of the underlying gossiping mechanism: whereas the earlier model assumed that an individual adjusts his opinion about somebody by gossiping about that person with everybody in the network, we assume instead that the individual gossips with that person about everybody. It turns out that the alternative model is able to lead to cooperative behaviour, unlike the previous model.

Citation: Traag VA, Van Dooren P, De Leenheer P (2013) Dynamical Models Explaining Social Balance and Evolution of Cooperation. PLoS ONE 8(4): e60063. doi:10.1371/journal.pone.0060063

Editor: Renaud Lambiotte, University of Namur, Belgium

Received September 28, 2012; **Accepted** February 25, 2013; **Published** April 25, 2013

Funding: This research received support from a grant "Actions de recherche concertées – Large Graphs and Networks" of the "Communauté Française de Belgique" and from the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization, http://sites.uclouvain.be/dysco/), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. PDL was provided a research fellowship by VLAC (Vlaams Academisch Centrum) and was hosted by Université catholique de Louvain as a visiting professor during sabbatical leave from the University of Florida, from which he received a Faculty Enhancement Opportunity. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Vincent.Traag@uclouvain.be

Introduction

Why do two antagonistic factions emerge so frequently in social networks? This question was already looming in the 1940s, when Heider [1] examined triads of individuals in networks, and postulated that only balanced triads are stable. A triad is balanced when friends agree in their opinion of a third party, while foes disagree, see Fig. 1. The individuals in an unbalanced triad have an incentive to adjust their opinions so as to reduce the stress experienced in such a situation [2]. Once an adjustment is made, the triad becomes balanced, and the stress disappears.

A decade later, Harary [3] showed that a complete social network splits in at most two factions if and only if all its triads are balanced, see also [4]. Such networks are called (socially) balanced as well. Since then, the focus of much of the research has been on detecting such factions in signed networks [5,6]. Many signed networks show evidence of social balance, although the split into factions might not be exact, that is, they are only nearly socially balanced [7–10].

What has been lacking until fairly recently, are dynamical models that explain *how* social balance emerges. The purpose of this paper is to analyse two such models. One of these models, proposed first in [11], was proved to exhibit social balance in [12]. However, this was done under a restrictive symmetry assumption for the reputation matrix. Here, we continue the analysis of this model and show that it generically does not lead to social balance when the symmetry assumption is dropped. In contrast, we propose a second model that is based on a different underlying gossiping mechanism, and show that it generically does lead to social balance, even when reputations are not symmetric.

Moreover, there is a natural connection between negative links and the evolution of cooperation: we consider positive links as indicating cooperation and negative links as defection. We will show that our alternative model is able to lead to cooperation, whereas the earlier model cannot.

Earlier Model

Certain discrete-time, stochastic dynamics have been investigated [13,14], but they exhibit so-called jammed states [15]: no change in the sign of a reputation improves the degree of social balance, as measured by the total number of balanced triads in the network. A surprisingly simple continuous-time model [11] was proved to converge to social balance for certain symmetric initial conditions [12]. The authors assume that the social network is described by a complete graph (everybody is connected to everybody), with weighted links representing reputations that change continuously in time. Let X denote the real-valued matrix of the reputations, so that X_{ij} represents the opinion i has about j. It is positive whenever i considers j a friend, and negative if i thinks of j as an enemy. The network is balanced, if, up to a possible relabelling of the individuals, the sign structure of X takes one of two possible block forms:

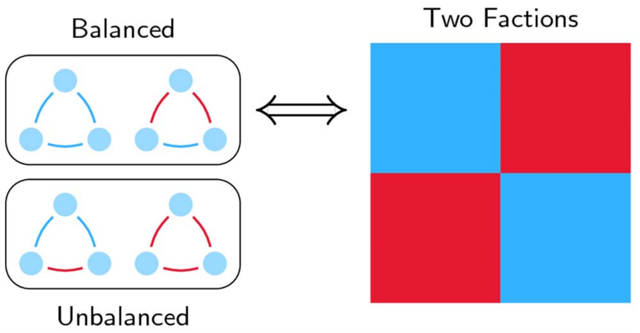

Figure 1. Social Balance. The two upper triads are balanced, while the two lower triads are unbalanced. According to the structure theorem [3], a complete graph can be split into (at most) two opposing factions, if and only if all triads are balanced. This is represented by the coloured matrix on the right, where blue indicates positive entries, and red negative entries.
doi:10.1371/journal.pone.0060063.g001

$$(+)\text{or}\begin{pmatrix} + & - \\ - & + \end{pmatrix}. \tag{1}$$

Changes in the reputations are modelled as follows:

$$\dot{X}=X^2,\ \text{or}\ \dot{X}_{ij}=\sum_k X_{ik}X_{kj}, \tag{2}$$

where $\dot{X}$ denotes the derivative with respect to time of the matrix X. The idea behind this model is that reputations are adjusted based on the outcome of a particular gossiping process. More specifically, suppose that Bob (individual i) wants to revise his opinion about John (individual j). Bob then asks everybody else in the network what they think of John. If one such opinion X_{kj} has the same sign as the opinion Bob has about his gossiping partner, i.e. as X_{ik}, then Bob will increase his opinion about John. But if these opinions differ in sign, then Bob will decrease his opinion about John.

The analysis for symmetric initial conditions $X(0)=X^T(0)$ was carried out in [12]: First, $X(0)$ is diagonalized by an orthogonal transformation $X(0)=U\Lambda(0)U^T$, where the columns of U are orthonormal eigenvectors $u_1,\ldots,u_n$ of $X(0)$ so that $UU^T=I_n$, and $\Lambda(0)$ is a diagonal matrix whose diagonal entries are the corresponding real eigenvalues $\lambda_1(0)\geq\lambda_2(0)\geq\cdot\geq\lambda_n(0)$ of $X(0)$. Direct substitution of the matrix function $U\Lambda(t)U^T$ shows that it is the solution of Eq. 2 with initial condition $X(0)$. Here, $\Lambda(t)$ is a diagonal matrix, solving the uncoupled matrix equation $\cdot\Lambda=\Lambda^2$ with initial condition $\Lambda(0)$. The diagonal entries of $\Lambda(t)$ are obtained by integrating the scalar first order equations $\dot{\lambda}_i=\lambda_i^2$:

$$\lambda_i(t)=\frac{\lambda_i(0)}{1-\lambda_i(0)t},\quad t\in\begin{cases}[0,+\infty)\text{if }\lambda_i(0)\leq 0\\ [0,1/\lambda_i(0))\text{if }\lambda_i(0)>0\end{cases} \tag{3}$$

Hence, the solution $X(t)$ blows up in finite time if and only if $\lambda_1(0)>0$. Moreover, if $\lambda_1(0)>0$ is a simple eigenvalue, then the solution $X(t)$, normalized by its Frobenius norm, satisfies:

$$\lim_{t\to 1/\lambda_1(0)}\frac{X(t)}{|X(t)|_F}=u_1u_1^T. \tag{4}$$

Assuming that u_1 has no zero entries, and up to a suitable permutation of its components, the latter limit takes one of the forms in Eq. 1. In other words, if the initial reputation matrix is symmetric and has a simple, positive eigenvalue, then the normalized reputation matrix becomes balanced in finite time.

Our first main result is that this conclusion remains valid for normal initial conditions, i.e. for initial conditions that satisfy the equality $X(0)X^T(0)=X^T(0)X(0)$, see SI Text S1, Theorem 2. Whereas the real eigenvalues behave similar to the symmetric case, the complex eigenvalues show circular behaviour, which results in small "bumps" in the dynamics as shown in Fig. 2 (see SI Fig. S1 for more detail). More precisely, if $X(0)$ is normal and if $\lambda_1(0)$ is a real, positive and simple eigenvalue which is larger than every other real eigenvalue (if any), then the solution $X(t)$ of Eq. 2 satisfies Eq. 4. Hence, once again, the normalized reputation matrix converges to a balanced state.

Our second main result is that this conclusion does not carry over to the case where $X(0)$ is not normal, see SI Text S1, Theorem 3. This general case is analysed by first transforming $X(0)$ into its real Jordan-canonical form $\mathcal{J}(0)$: $X(0)=T\mathcal{J}(0)T^{-1}$, where T consists of a basis of (the real and imaginary parts of) generalized eigenvectors of $X(0)$. It can then be shown that the solution $X(t)$ of Eq. 2 is given by $T\mathcal{J}(t)T^{-1}$, where $\mathcal{J}(t)$ solves the matrix equation $\dot{J}=J^2$, an equation which can still be solved explicitly. Hence, $X(t)$ can still be determined. It turns out that if $X(0)$ has a real, positive and simple eigenvalue $\lambda_1(0)$ which is larger than every other real eigenvalue (if any), then the normalized reputation matrix satisfies:

$$\lim_{t\to 1/\lambda_1(0)}\frac{X(t)}{|X(t)|_F}=\frac{u_1v_1^T}{|u_1v_1^T|_F}, \tag{5}$$

where u_1 and v_1^T are left and right eigenvectors of $X(0)$ respectively, that correspond to the eigenvalue $\lambda_1(0)$. If we assume that none of the entries of u_1 and v_1 are zero, then we can always find a suitable permutation of the components of u_1 and v_1 such that they have the following sign structure:

$$u_1=\begin{pmatrix}+\\+\\\hline-\\-\end{pmatrix}\text{and } v_1^T=(+\quad-\mid+\quad-)$$

Consequently, in general, the matrix limit in Eq. 5 has the sign structure:

$$\left(\begin{array}{cc|cc}+ & - & + & -\\ - & + & - & +\end{array}\right),$$

as illustrated in Fig. 2. Clearly, this configuration doesn't correspond to social balance any longer.

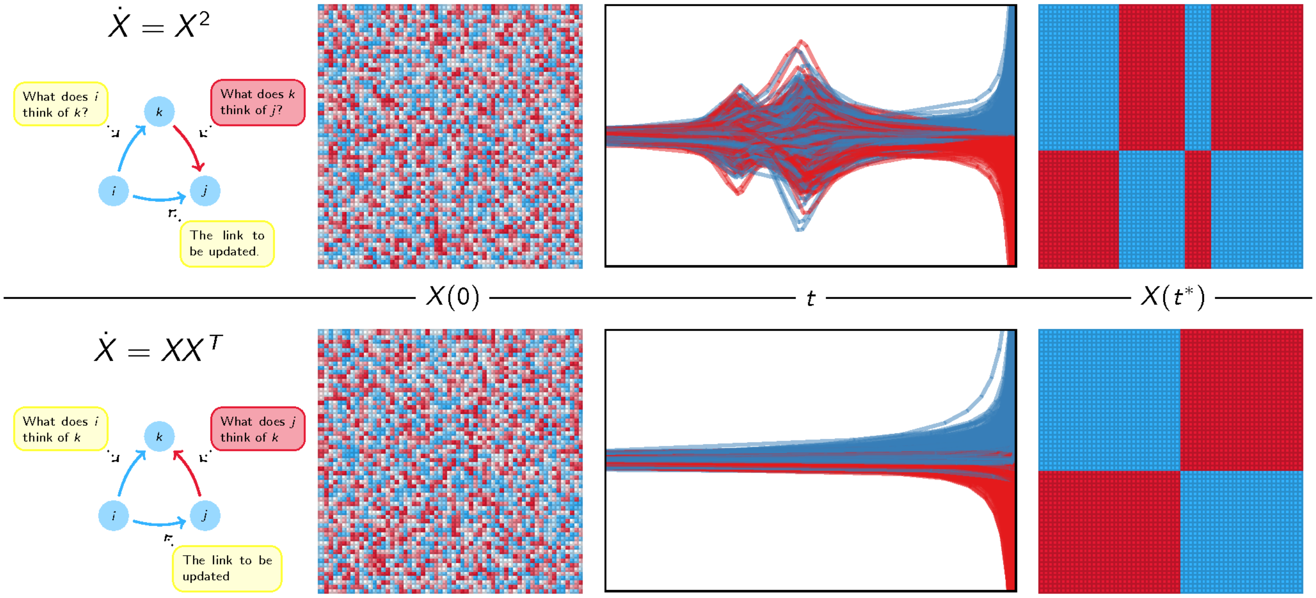

Figure 2. The two models compared. The first row illustrates what happens generically for the model $\dot{X}=X^2$, while the second row displays the results for $\dot{X}=XX^T$. Each row contains from left to right: (1) an illustration of the model; (2) the random initial state; (3) the dynamics of the model; and (4) the final state to which the dynamics converge. Blue indicates positive entries, and red negative entries. Although the first model converges to a rank one matrix, it is not socially balanced. The second model does converge generically to social balance. The small bumps in the dynamics for $\dot{X}=X^2$ are due to complex eigenvalues that show circular behaviour (see Fig. S1).
doi:10.1371/journal.pone.0060063.g002

Alternative Model

Let us briefly reconsider the gossiping process underlying model $\dot{X}=X^2$. In our example of Bob and John, the following happens. Bob asks others what they think of John. Bob takes into account what he thinks of the people he talks to, and adjusts his opinion of John accordingly. An alternative approach is to consider a type of homophily process [16–18]: people tend to befriend people who think alike. When Bob seeks to revise his opinion of John, he talks to John about everybody else (instead of talking to everybody else about John). For example, suppose that Bob likes Alice, but that John dislikes her. When Bob and John talk about Alice, they notice they have opposing views about her, and as a result the relationship between Bob and John deteriorates. On the other hand, should they share similar opinions about Alice, their relationship will improve. Thus, our alternative model for the update law of the reputations is:

$$\dot{X}=XX^T, \text{ or } \dot{X}_{ij}=\sum_k X_{ik}X_{jk}. \tag{6}$$

Although there apparently is only a subtle difference in the gossiping processes underlying the models in Eq. 2 and 6, these models turn out to behave quite differently, as we discuss next.

Our third main result is that for generic initial conditions, the normalized solution of system Eq. 6 converges to a socially balanced state in finite time. To show this, we decompose the solution $X(t)$ into its symmetric and skew-symmetric parts: $X(t)=S(t)+A(t)$, where $S(t)=S^T(t)$ and $A(t)=-A^T(t)$. Since $\dot{X}=\dot{X}^T$, the skew-symmetric part remains constant, and therefore $A(t)=A(0) \equiv A_0$. The symmetric part then obeys the matrix Riccati differential equation $\dot{S}=(S+A_0)(S-A_0)$. We introduce $Z(t)=e^{-A_0 t}S(t)e^{A_0 t}$ to eliminate the linear terms in this equation, and obtain

$$\dot{Z}=Z^2+A_0A_0^T. \tag{7}$$

The latter matrix Riccati differential equation can be integrated, yielding the solution $Z(t)$ explicitly, and hence $S(t)$, as well as $X(t)$, can be calculated.

It turns out that if $A_0 \neq 0$, then $X(t)$ always blows up in finite time. Moreover, using a perturbation argument, it can be shown there is a dense set of initial conditions $X(0)$ such that the normalized solution of Eq. 6 converges to

$$\lim_{t\to t*}\frac{X(t)}{|X(t)|_F}=ww^T, \tag{8}$$

for some vector w, as t approaches the blow-up time t*, see SI Text S1, Theorem 5. If w has no zero entries, this implies that the normalized solution becomes balanced in finite time. Hence, the alternative model in Eq. 6 generically evolves to social balance, see Fig. 2.

Evolution of Cooperation

Positive and negative links have a natural interpretation in the light of cooperation: positive links indicate cooperation and negative links indicate defection. There is then also a natural motivation for the alternative model in terms of cooperation. Again, suppose Bob wants to revise his opinion of John. For Bob it is important to know whether John is cooperative in order to determine whether he should cooperate with John or not. So, instead of asking Alice whether she has cooperated with John, Bob would like to know whether John has cooperated with her. In other words, Bob is not interested in X_{kj} but in X_{jk}, consistent with

Eq. 6, illustrated in Fig. 2. This is also what is observed in studies on gossip: it often concerns what others did, not what one thinks of others [19,20].

Indeed gossiping seems crucial in explaining the evolution of human cooperation through indirect reciprocity [21]. It has even been suggested that humans developed larger brains in order to gossip, so as to control the problem of cooperation through social interaction [22]. In general, the problem is that if defection allows individuals to gain more, why then do individuals cooperate? This is usually modelled in the form of a prisoner's dilemma, in which each agent has the possibility to give his partner some benefit b at some cost $c<b$. So, if an agent's partner cooperates (he gives the agent b) but the agent doesn't cooperate (he doesn't pay the cost c) his total payoff will be b. Considering the other possibilities results in the payoff matrix detailed in Fig. 3.

Irrespective of the choice of the other player, it is better to defect in a single game. Suppose that the second player cooperates. Then if the first player cooperates he gains $b-c$, while if he defects he gains b, so defecting is preferable. Now suppose that the second player defects. The first player then has to pay c, but doesn't have to pay anything when defecting. So indeed, in a single game, it is always better to defect, yet the payoff is higher if both cooperate, whence the dilemma.

In reality, we do observe cooperation, and various mechanisms for explaining the evolution of cooperation have been suggested [23], such as kin selection [24,25], reciprocity [26] or group selection [27]. Humans have a tendency however to also cooperate in contexts beyond kin, group or repeated interactions. It is believed that some form of indirect reciprocity can explain the breadth of human cooperation [21]. Whereas in direct reciprocity the favour is returned by the interaction partner, in indirect reciprocity the favour is returned by somebody else, which usually involves some reputation. It has been theorized that such a mechanism could even form the basis of morality [28]. Additionally, reputation (and the fear of losing reputation) seems to play an important role in maintaining social norms [29–31].

In general, the idea is the following: agents obtain some good reputation by helping others, and others help those with a good reputation. Initially a strategy known as image scoring was introduced [32]. Shortly after, it was argued that a different strategy, known as the standing strategy, should actually perform better [33], although experiments showed people tend to prefer the simpler image scoring strategy [34]. This led to more systematic studies of how different reputation schemes would perform [35–37]. Although much research has been done on indirect reciprocity, only few theoretical works actually study how gossiping shapes reputations [38,39]. Nonetheless, most studies (tacitly) assume that reputations are shaped through gossip. Additionally, it was observed experimentally that gossiping is an effective mechanism for promoting cooperation [40–42].

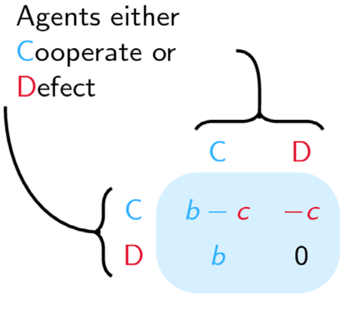

Figure 3. Prisoner's Dilemma. Both players have the option to either Cooperate or Defect. Whenever an agent cooperates, it costs him c while his partners receives a benefit $b>c$, leading to the indicated payoffs.
doi:10.1371/journal.pone.0060063.g003

Moreover, these reputations are usually considered as objective. That is, all agents know the reputation X_j of some agent j, and all agents have the same view of agent j. Private reputations–so that we have X_{ij}, the reputation of j in the eyes of i–have usually been considered by allowing a part of the population to "observe" an interaction, and update the reputation accordingly. If too few agents are allowed to "observe" an interaction, the reputations X_{ij} tend to become uncorrelated and incoherent. This makes reputation unreliable for deciding whether to cooperate or defect. The central question thus becomes how to model private reputations such that they remain coherent and reliable for deciding whether to cooperate or not.

Dynamical models of social balance might provide an answer to this question. Although it allows to have private reputations–that is X_{ij}–the dynamics could also lead to some coherence in the form of social balance. In addition, it models more explicitly the gossiping process, commonly suggested to be the foundation upon which reputations are forged.

Simulation Results

The reputations of the agents are determined by the dynamics of the two models. We call agents using $\dot{X}=X^2$ dynamics type A, and those using $\dot{X}=XX^T$ dynamics type B. We assume that agent i cooperates with j whenever $X_{ij}>0$ and defects otherwise. Agents reproduce proportional to their fitness, determined by their payoff. Agents that do well (have a high payoff) have a higher chance of reproduction, and we are interested in knowing the probability that a certain type becomes fixated in the population (i.e. takes over the whole population), known as the fixation probability ρ. All simulations start off with an equal amount of agents, so if some type wins more often than his initial relative frequency, it indicates it has an evolutionary advantage. For the results presented here this comes down to $\rho>1/2$. More details on the simulations are provided in the Materials and Methods section at the end of the paper.

The results are displayed in Fig. 4 using a normalized cost of $c=1$ (the ratio b/c drives the evolutionary dynamics, see Materials and Methods and [23]). When directly competing against each other, type B has an evolutionary advantage (its fixation probability $\rho_B>1/2$) compared to type A, already for relatively small benefits. When each type is playing against defectors (agents that always defect), type A seems unable to defeat defectors ($\rho_A<1/2$) for any $b<20$, while type B performs quite well against them. When all three types are playing against each other results are similar (see SI Fig. S2). When varying the number of agents, the critical benefit b^* at which type B starts to have an evolutionary advantage changes (i.e. where the fixation probability $\rho_B=1/2$). For $b>b^*$ agents using the model $\dot{X}=XX^T$ have a higher chance to become fixated, while for $b<b^*$ defectors tend to win. The inequality for type B to have an evolutionary advantage can be relatively accurately approximated by $b>b*=\gamma\sqrt{n}$ where γ is estimated to be around $\gamma\approx1.72\pm0.037$ (95% confidence interval). Varying the intensity of selection does not alter the results qualitatively (see SI Fig. S3). Summarizing, type B is able to lead to cooperation and defeats type A. Based on these results, if a gossiping process evolved during the course of human history in order to maintain cooperation, the model $\dot{X}=XX^T$ seems more likely to have evolved than $\dot{X}=X^2$. For smaller groups a smaller benefit is needed for the model $\dot{X}=XX^T$ to become fixated. This

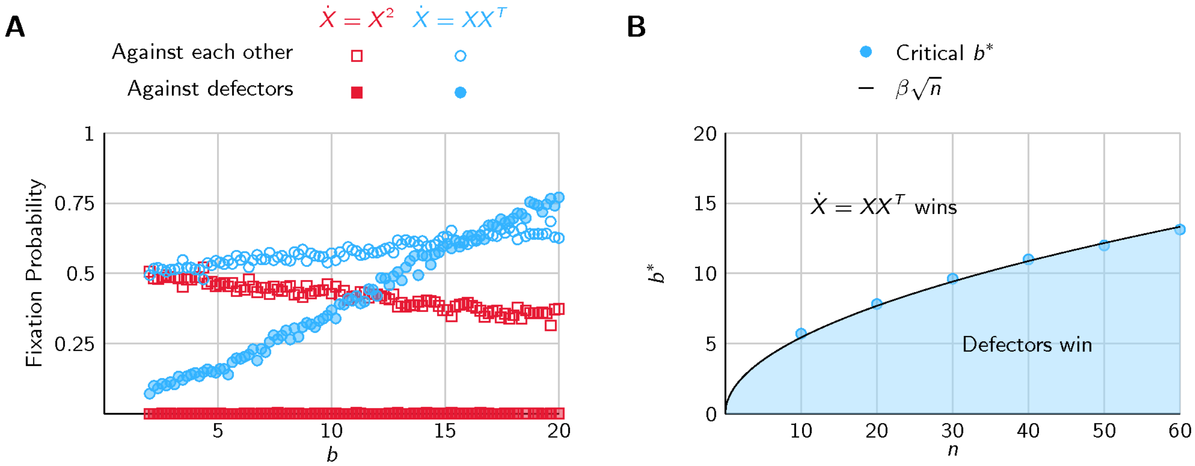

Figure 4. Evolution of Cooperation. (A) The fixation probability (probability to be the sole surviving species) is higher for model $\dot{X}=XX^T$ than $\dot{X}=X^2$. This implies that the model $\dot{X}=XX^T$ is more viable against defectors, and has an evolutionary advantage compared to $\dot{X}=X^2$. (B) The point b^* at which the model $\dot{X}=XX^T$ has an evolutionary advantage against defectors (i.e. the fixation probability $\rho > 1/2$) depends on the number of agents n. The condition for the model $\dot{X}=XX^T$ to defeat defectors can be approximated by $b>b*=\beta\sqrt{n}$, with $\beta\approx 1.72$.
doi:10.1371/journal.pone.0060063.g004

dependence seems to scale only as $\sqrt{n}$, so that larger groups only need a marginally larger benefit in order to develop cooperation.

Conclusion

To conclude, we have shown that the alternative model $\dot{X}=XX^T$ generically converges to social balance, whereas the model $\dot{X}=X^2$ did not. The current models exhibit several unrealistic features, we would like to address: (1) an all-to-all topology; (2) dynamics that blow-up in finite time; and (3) homogeneity of all agents. Although most of these issues can be addressed by specifying different dynamics, the resulting models are much more difficult to analyse, thereby limiting our understanding. Although the two models are somewhat simple, they are also tractable, and what we lose in truthfulness, we gain in deeper insights: in simplicity lies progress. Our current analysis offers a quite complete understanding, and we hope it provides a stepping stone to more realistic models, which we would like to analyse in the future.

The difference between the two models can be understood in terms of gossiping: we assume that people who wish to revise their opinion about someone talk to that person about everybody else, while the earlier model assumed that people talk about that person to everybody else. Both gossiping and social balance are at the centre of many social phenomena [22,29,43,44], such as norm maintenance [30], stereotype formation [45] and social conflict [46]. For example, a classic work [29] on the established and outsiders found that gossiping was the fundamental driving force for the maintenance of the cohesive network of the established at the exclusion of the outsiders. Understanding how social balance may emerge might help to understand the intricacies of these social phenomena.

Moreover, in light of the evolution of cooperation it appears that agents using $\dot{X}=XX^T$ dynamics perform well against defectors, and have an evolutionary advantage compared to agents using $\dot{X}=X^2$ dynamics. Contrary to other models of indirect reciprocity, not everybody might end up cooperating with everybody, and the population may split into two groups. This provides an interesting connection between social balance theory, gossiping and the evolution of cooperation. Our results improve our understanding of gossiping as a mechanism for group formation and cooperation, and as such contributes to the study of indirect reciprocity.

Materials and Methods

In the simulations of the evolution of cooperation, the dynamics consist of two parts: (1) the interaction dynamics within each generation; and (2) the dynamics prescribing how the population evolves from generation to generation.

Interaction Dynamics

We include three possible types of agents in our simulations:
Type A uses $\dot{X}=X^2$ dynamics,
Type B uses $\dot{X}=XX^T$ dynamics, and
Defectors have trivial reputation dynamics $\dot{X}=0$, with negative constant reputations.

We can decompose the reputation matrix $X(t)$ accordingly into three parts:

$$X(t)=\begin{pmatrix} X_A(t) \\ X_B(t) \\ X_D(t) \end{pmatrix},$$

where $X_A(t)$ are the reputations of all agents in the eyes of agents of type A, $X_B(t)$ for type B and $X_D(t)$ for defectors. The reputations $X_A(0)$ and $X_B(0)$ are initialized from a standard Gaussian distribution. The initial reputation for $X_D(0)$ will be set to a fixed negative value. To be clear, $X_D(0)$ is the reputation of all other agents in the eyes of defectors, which is negative initially. The initial reputation of the defectors themselves is of course not necessarily negative initially. For the results displayed here we have used $X_D(0)=-10$, but results remain by and large the same when varying this parameter, as long as it remains sufficiently negative.

Since we are dealing with continuous dynamics in this paper, we assume all agents are involved in infinitesimally short games at each time instance t. Each agent i may choose to either cooperate or defect with another agent j, and this decision may vary from one agent to the next. For agents of type A and type B the decision to cooperate is based on the reputation: they defect whenever $X_{ij}(t) \leq 0$ and cooperate whenever $X_{ij}(t) > 0$. We define the cooperation matrix $C(t)$ accordingly.

$$C_{ij}(t) = \begin{cases} 0 & \text{if } X_{ij} \leq 0 \\ 1 & \text{if } X_{ij} > 0 \end{cases}$$

Defectors will simply always defect. Whenever an agent i cooperates with j the latter receives a payoff of b at a cost of c to agent i. We integrate this payoff over all infinitesimally short games from time 0 to time t^*, which can be represented as

$$p(g) = \frac{1}{n}\int_0^{t*} bC(t)^T e - cC(t)e dt,$$

where $e = (1,\ldots,1)$ the vector of all ones for a certain generation g.

Evolutionary Dynamics

We have simulated the evolution of cooperation for $n = 10,20,\ldots,60$ agents, which stays constant throughout evolution. We consider four different schemes for initializing the first generation:

	$p_A(0)$	$p_B(0)$	$p_D(0)$
1. Type A vs Type B	1/2	1/2	-
2. Type A vs Defectors	1/2	-	1/2
3. Type B vs Defectors	-	1/2	1/2
4. Type A,B and Defectors	1/3	1/3	1/3

Here $p_A(0)$,$p_B(0)$ and $p_D(0)$ are respectively the proportion of agents of type A, type B and defectors in the first generation. We use the vector $T_i(g) \in \{A, B, D\}$ to denote the type of agent i in generation g, so that $T_i(g) = A$ if agent i is a type A player, $T_i(g) = B$ for a type B player, and $T_i(g) = D$ for a defector. We are interested in estimating the probability that a single type takes over the whole population, known as the fixation probability ρ_A, ρ_B and ρ_D for the three different types. If a type has no evolutionary advantage, it is said to be evolutionary neutral, and in that case its fixation probability is equal to its initial frequency, e.g. for type A $\rho_A = p_A(0)$.

We will keep the population constant at the initial n, and simply choose n new agents according to their payoff for the next generation. This can be thought of as choosing n times one of the n agents in the old generation for reproduction. Let ϕ_i denote the probability that an agent is selected for reproduction, which we define as

$$\phi_i = \frac{e^{\beta P_i(g)}}{\sum_i e^{\beta P_i(g)}}.$$

Since we are only interested in the number of agents of a certain type, we can also gather all payoffs for the same type of agents, and write

$$\Phi_q = \sum_{i:T_i(g)=q} \phi_i,$$

where $q \in \{A, B, D\}$ represents the type of agent. The probability to select a type A agent, a type B agent or a defector is then respectively Φ_A, Φ_B and Φ_D. In the next generation, the probability that agent i is of a specific type q can then be written as

$$\Pr(T_i(g+1) = q) = \Phi_q.$$

This evolutionary mechanism can be seen as a Wright-Fisher process [47] with fitnesses $e^{\beta P_i(g)}$. It is well known that this process converges faster than a Moran birth-death process, since it essentially takes n time steps in a Moran process to reproduce the effect of one time step in a Wright-Fisher process [47]. Because of the high computational costs (solving repeatedly a non-linear system of differential equations of size n^2), this process is preferable.

Higher β signifies higher selective pressure, and leads to a higher reproduction of those with a high payoff, and in the case that $\beta \to \infty$ only those with the maximum payoff reproduce. On the other hand, for $\beta \to 0$ this tends to the uniform distribution $\phi_i = 1/n$, where payoffs no longer play any role. We have used $\beta = 0.5$ for the low selective pressure, $\beta = 5$ for the high selective pressure, reported in SI Fig. S3. For the results in the main text we have used $\beta = 1$.

For an evolutionary neutral selection in where all $P_i(g) = P$ are effectively the same, β has no effect, and $\phi_i = 1/n$. Notice that if we rescale $P_i(g)$ by $1/c$ so that the payoff effectively becomes

$$\frac{1}{c}P_i(g) = \frac{1}{n}\int_0^{t*} \frac{b}{c}C(t)^T e - C(t)e dt,$$

and we rescale β by c, then the reproduction probabilities remain unchanged. Hence, only the ratio b/c effectively plays a role up to a rescaling of the intensity of selection. Since the point at which the evolution is neutral (i.e. ρ equals the initial proportional frequency), is independent of β, this point will only depend on the ratio b/c. So, we normalized the cost $c = 1$. To verify this, we also ran additional simulations with different costs, which indeed gave the same results.

We stop the simulation whenever one of the types becomes fixated in the population. With *fixation* we mean that all other types have gone extinct, and only a single type remains. If no type has become fixated after 1,000 generations, we terminate the simulation and count as winner the most frequent type. This almost never happens, and the simulation usually stops after a relatively small number of generations.

In total, we repeat this process 1,000 times for the results in the main text, and for the low ($\beta = 0.5$) and high ($\beta = 5$) selective pressure 100 times. This means that we run the evolutionary dynamics until one of the types has become fixated, and we record which type has "won". After that, we again start from the first generation, and run until fixation, and repeat this. Finally, we calculate how many rounds a type has "won" compared to the total number of rounds, which yields the fixation probability ρ.

Supporting Information

Figure S1 Phase portrait of system S12-S13. Circular orbits in the upper half plane ($a > 0$) are traversed counter clockwise, whereas circular orbits in the lower half plane ($a < 0$) are traversed clockwise.
(TIFF)

Figure S2 Results including type A, B and defectors.
(TIFF)

Figure S3 Results different intensities of selection.
(TIFF)

Text S1 Proofs and details of statements in the main paper.
(PDF)

Author Contributions

Conceived and designed the experiments: VAT. Performed the experiments: VAT. Analyzed the data: VAT PVD PDL. Contributed reagents/materials/analysis tools: VAT PVD PDL. Wrote the paper: VAT PVD PDL.

References

1. Heider F (1946) Attitudes and Cognitive Organization. J Psychol 21: 107–112.
2. Bearman PS, Moody J (2004) Suicide and Friendships Among American Adolescents. Am J Public Health 94: 89–95.
3. Harary F (1953) On the notion of balance of a signed graph. The Michigan Mathematical Journal 2: 143–146.
4. Cartwright D, Harary F (1956) Structural balance: a generalization of Heider's theory. Psychol Rev 63: 277–293.
5. Doreian P, Mrvar A (1996) A partitioning approach to structural balance. Soc Networks 18: 149–168.
6. Traag VA, Bruggeman J (2009) Community detection in networks with positive and negative links. Phys Rev E 80: 036115.
7. Szell M, Lambiotte R, Thurner S (2010) Multirelational organization of large-scale social networks in an online world. Proc Natl Acad Sci U S A 107: 13636–41.
8. Leskovec J, Huttenlocher D, Kleinberg J (2010) Predicting positive and negative links in online social networks. In: WWW 2010.
9. Facchetti G, Iacono G, Altafini C (2011) Computing global structural balance in large-scale signed social networks. Proc Natl Acad Sci U S A 108: 20953–20958.
10. Kunegis J, Lommatzsch A, Bauckhage C (2009) The slashdot zoo. In: Proceedings of the 18th international conference on World wide web - WWW '09. New York, New York, USA: ACM Press, p. 741.
11. Kulakowski K, Gawronski P, Gronek P (2005) The Heider balance - a continuous approach. Int J Mod Phys C 16: 707–716.
12. Marvel SA, Kleinberg J, Kleinberg RD, Strogatz SH (2011) Continuous-time model of structural balance. Proc Natl Acad Sci U S A 108: 1771–6.
13. Antal T, Krapivsky PL, Redner S (2005) Dynamics of social balance on networks. Phys Rev E 72: 36121.
14. Radicchi F, Vilone D, Yoon S, Meyer-Ortmanns H (2007) Social balance as a satisfiability problem of computer science. Phys Rev E 75: 026106.
15. Marvel S, Strogatz S, Kleinberg J (2009) Energy Landscape of Social Balance. Phys Rev Lett 103: 198701.
16. Mcpherson M, Smith-Lovin L, Cook JM (2001) Birds of a Feather: Homophily in Social Networks. Annu Rev Sociol 27: 415–444.
17. Durrett R, Levin SA (2005) Can stable social groups be maintained by homophilous imitation alone? J Econ Behav Organ 57: 267–286.
18. Fu F, Nowak MA, Christakis NA, Fowler JH (2012) The Evolution of Homophily. Scientific Reports 2.
19. McAndrew FT, Bell EK, Garcia CM (2007) Who Do We Tell and Whom Do We Tell On? Gossip as a Strategy for Status Enhancement. J Appl Soc Psychol 37: 1562–1577.
20. Paine R (1967) What is Gossip About? An Alternative Hypothesis. Man 2: 278–285.
21. Nowak MA, Sigmund K (2005) Evolution of indirect reciprocity. Nature 437: 1291–1298.
22. Dunbar RIM (1998) Grooming, Gossip, and the Evolution of Language. Cambridge: Harvard University Press, 242 pp.
23. Nowak MA (2006) Five rules for the evolution of cooperation. Science (New York, NY) 314: 1560–3.
24. Smith JM, Maynard Smith J (1982) Evolution and the Theory of Games. Cambridge: Cambridge University Press, 226 pp.
25. Hamilton W (1964) The genetical evolution of social behaviour I. J Theor Biol 7: 1–16.
26. Axelrod R, Hamilton W (1985) The Evolution of Cooperation, volume 211. New York: Basic Books, 1390–1396 pp.
27. Wilson DS (1975) A theory of group selection. Proc Natl Acad Sci U S A 72: 143–146.
28. Alexander RD (1987) The Biology of Moral Systems. New York: Aldine de Gruyter.
29. Elias N, Scotson JL (1994) The Established and the Outsiders. London: SAGE Publications.
30. Friedkin NE (2001) Norm formation in social influence networks. Soc Networks 23: 167–189.
31. Fehr E, Fischbacher U (2004) Third-party punishment and social norms. Evolution and Human Behavior 25: 63–87.
32. Nowak MA, Sigmund K (1998) Evolution of indirect reciprocity by image scoring. Nature 393: 573–7.
33. Leimar O, Hammerstein P (2001) Evolution of cooperation through indirect reciprocity. Proc Biol Sci 268: 745–753.
34. Milinski M, Semmann D, Bakker TC, Krambeck HJ (2001) Cooperation through indirect reciprocity: image scoring or standing strategy? Proc Biol Sci 268: 2495–2501.
35. Ohtsuki H, Iwasa Y (2006) The leading eight: social norms that can maintain cooperation by indirect reciprocity. J Theor Biol 239: 435–44.
36. Brandt H, Sigmund K (2004) The logic of reprobation: assessment and action rules for indirect reciprocation. J Theor Biol 231: 475–486.
37. Ohtsuki H, Iwasa Y (2004) How should we define goodness?–reputation dynamics in indirect reciprocity. J Theor Biol 231: 107–120.
38. Nakamaru M, Kawata M (2004) Evolution of rumours that discriminate lying defectors. Evol Ecol Res 6: 261–283.
39. Traag VA, Van Dooren P, Nesterov Y (2011) Indirect reciprocity through gossiping can lead to cooperative clusters. In: IEEE Symposium on Artificial Life 2011. Piscataway: IEEE, pp.154–161.
40. Piazza J, Bering JM (2008) Concerns about reputation via gossip promote generous allocations in an economic game. Evolution and Human Behavior 29: 172–178.
41. Sommerfeld RD, Krambeck HJ, Milinski M (2008) Multiple gossip statements and their effect on reputation and trustworthiness. Proc Biol Sci 275: 2529–2536.
42. Sommerfeld RD, Krambeck HJ, Semmann D, Milinski M (2007) Gossip as an alternative for direct observation in games of indirect reciprocity. Proc Natl Acad Sci U S A 104: 17435–17440.
43. Gluckman M (1963) Gossip and Scandal. Curr Anthropol 4: 307–316.
44. Foster EK (2004) Research on gossip: Taxonomy, methods, and future directions. Rev Gen Psychol 8: 78–99.
45. Wert SR, Salovey P (2004) A social comparison account of gossip. Rev Gen Psychol 8: 122–137.
46. Labianca G, Brass D, Gray B (1998) Social networks and perceptions of intergroup conflict: The role of negative relationships and third parties. Academy of Management journal 41: 55–67.
47. Gillespie JH (2004) Population Genetics: A Concise Guide. Baltimore: The John Hopkins University Press.