Emergence of Cooperation in Anonymous Social Networks through Social Capital*

Nicole Immorlica
Northwestern University
Evanston, IL USA
nicimm@gmail.com

Brendan Lucier†
University of Toronto
Toronto, ON Canada
blucier@cs.toronto.edu

Brian Rogers
Northwestern University
Evanston, IL USA
b-rogers@kellogg.northwestern.edu

ABSTRACT

We study the emergence of cooperation in dynamic, anonymous social networks, such as in online communities. We examine prisoner's dilemma played under a social matching protocol, where individuals form random links to partners with whom they can interact. Cooperation results in mutual benefits, whereas defection results in a high short-term gain. Moreover, an agent that defects can escape reciprocity by virtue of anonymity: it is always possible for an agent to abandon his history and re-enter the network as a new user. We find that cooperation is sustainable at equilibrium in such a model. Indeed, cooperation allows an individual to interact with an increasing number of other cooperators, resulting in the formation of a type of social capital. This process arises endogenously, without the need for potentially harmful social enforcement rules. Additionally, for a rich class of parameter settings, our model predicts a stable coexistence of cooperating and defecting agents at equilibrium.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*

General Terms

Networks, Interaction Models

Keywords

Social Networks, Prisoner's Dilemma, Equilibria

1. INTRODUCTION

In many important social, political, and economic situations, people face a choice between seeking immediate personal gain at the expense of others or cooperating to a lesser mutual benefit. In such a scenario, it seems at first glance that an agent acting in his own best interest ought to choose an uncooperative strategy. This intuition is amplified when people can act under pseudonyms, such as over the Internet, since an agent that develops a reputation for being uncooperative can simply re-enter the system with a new name. However, the continuing success of online interaction networks such as eBay (www.ebay.com) indicates that a group of anonymous agents need not devolve into a steady state of completely non-cooperative behaviour. We find that such cooperation can be explained as an equilibrium of rational behaviour in a simple network formation model.

We model the choice between mutual and personal gain by the classic 2-player Prisoner's Dilemma. In this game, each player has the option of cooperation or defection. If both players cooperate then they both receive modest payoffs, but if one player defects then he will receive a higher utility while the cooperator suffers a penalty. If both agents defect then they both receive no payoff. It is not hard to see that each player acting in his own self interest ought to defect – no matter what the action of his opponent, a player maximizes his payoff by defecting. In game-theoretic terms, we say defecting is a *dominant strategy* and thus provides a strong prediction of play in such games.

Examples abound, both in field observations and laboratory experiments, in which participants choose to cooperate in the prisoner's dilemma. A common explanation is that reputations can influence behaviour: a history of defection will follow an agent and be visible to potential partners, affecting future payoffs. In practice, cooperation is encouraged in online networks through the implementation of exogenous reputation systems that assist in this flow of information. For example, eBayallows buyers and sellers to make publicly-visible comments about each other, question-and-answer site AllExperts (www.allexperts.com) includes a ranking system for experts, and gaming sites such as the Internet Go Server (www.pandanet.co.jp/English/) keep records of game history that can be used to verify player skill level. In each case, a deviating agent can expect repercussions in future interactions. However, when agents can shed reputations by virtue of changing names at no cost, these systems seem insufficient by themselves to explain cooperative behaviour.

Early research into games on social networks attempted to provide models supporting cooperation in the absence of reputation mechanisms. When the partners in the game are fixed over time and the game is repeated infinitely, a classic application of the Folk Theorem provides a vehicle for

*We thank Wiola Dziuda and Christoph Kuzmics for helpful comments.

†The work was performed in part while this author was visiting Northwestern University.

EC 2010 Boston, MA USA

cooperation (or any mutually beneficial payoff). Namely, agents can use the threat of defection ("if you defect, I'll defect forever") to sustain cooperation [17]. When agents change partners over time, such threats are no longer sustainable because a pair of agents may very well never meet again. Instead, community enforcement procedures are used to sustain cooperation [25, 29]. In these equilibria, if the model supports public reputations, then the community can always defect against an agent with a reputation of defection. These explanations for cooperation are unsatisfactory in the domain of anonymous networks: threat of retaliation is not a deterrant when an agent can re-enter the network as a new user at any time.

An alternative explanation, for the case that agents are anonymous and plays are not publicly observable, is that the community can enforce cooperation by agreeing to defect with all partners as soon as any defection is observed [12]. In this way, a deviating defector starts a defection contagion and will thus eventually be punished for the initial defection. Such a system is highly unstable, as the presence of even a single defecting agent would preclude any hope of a cooperative state. Since real-world examples generally feature at least some degree of non-cooperation, it seems that this is again insufficient to predict observed behaviour.

In the above models, partnerships are chosen exogenously, be they fixed or random. More recent literature questions the effect of allowing agents to *choose* partners to a certain extent [9, 18, 26, 34]. We employ such an approach: agent behaviour will influence network formation. Our model consists of fully anonymous agents, who build social capital implicitly in the form of a neighbourhood of partners. Unlike other related work, our model sustains, in equilibrium, a heterogeneous society in which there are non-trivial fractions of both cooperating and defecting agents. In many cases such equilibria are stable, resilient to deviating play by a small number of agents. Thus, the presence of (or an increase in) non-cooperative play need not be detrimental to the long-term success of the network.

1.1 The Model

In our model, there is a countable set of agents interacting with each other on a directed network. The network is dynamic, with agents entering and leaving over time. Nodes are removed from the network at a fixed uniform rate, and whenever an agent is removed it is replaced by a new agent.

Each agent sponsors a bounded number of connections to other agents. Those relationships that an agent sponsers are termed its outlinks; its relationships sponsered by others are its inlinks. Each connection persists through time until one of the partners dies or chooses to break it. When a connection is broken, the agent who sponsored it randomly re-matches with another agent at the next time period. On each round, an agent plays a prisoner's dilemma game with each of his (incoming and outgoing) neighbours, choosing the same strategy for each neighbour.

Our model abstracts an interaction network in which a reputation system enables perfect information flow, but the agents have anonymity through the use of pseudonyms. A node is meant to represent an agent with a given pseudonym; we assume that each agent uses only one pseudonym at each timestep.[1] Given the perfect information flow from the reputation system, an agent that defects in some interactions would quickly be identified as untrustworthy, and not interacted with by any other agents; at this point his only recourse would be to re-enter the system with a new pseudonym. This motivates our assumption, standard in the study of non-cooperative games on social networks, that an agent plays the same strategy in all interactions each round: an agent who wishes to defect should defect in all interactions, as he would likely lose all links regardless.

How will rational agents choose their strategies in this model? We begin our analysis by making three simplifying assumptions about agent behaviour. The first assumption is that agents are *unforgiving*; agents sever relationships with defectors in favor of new random partners (and always maintain relationships with cooperators). Note that breaking all links to and from a node is equivalent to having that node leave and re-enter the network; this corresponds to our intuition about anonymous agents being able to start fresh when discovered as non-cooperative. The second assumption is that agents are *consistent* in their behavior: agents commit to a strategy at birth and never change. The third assumption is that agents are *trusting* of strangers: they always accept proposed inlinks.

Under these assumptions, we show that our model supports cooperation in the society. Since relationships to defectors are broken, an agent can only build up a network of relationships by committing to cooperation. The rate at which this happens is a (nonlinear) function of the fraction of cooperators and defectors in the society. These inlinks become a valuable asset, which can be thought of as the social capital of the agent. If the expected lifetime of an agent is sufficiently high, the promise of this asset may induce selfish agents to commit to cooperation even though the per-period-per-interaction payoff for cooperation is lower than that of defection. When the expected lifetime utility of defecting equals that of cooperating, the model supports in equilibrium a heterogeneous society in which cooperators and defectors co-exist. We also show that some of these equilibria are stable in that, should the fraction of cooperators grow (or shrink) beyond the equilibrium mixture, then newborn agents will prefer to defect (or cooperate) and thus self-correct the composition. We emphasize that, unlike other work, the presence of defectors in our model arises naturally due to the strategic choices of defection and cooperation, and is not an assumption in our model.

We now briefly describe ways in which our three assumptions (unforgiving, consistent, and trusting agents) can be motivated or relaxed.[2] First, if agents are consistent, then it is equilibrium behavior for them to be unforgiving (since defection in a partner indicates that no further cooperative benefit can be gained from the interaction). Even if agents are not assumed to be consistent, we feel that it is reasonable to assume that they are unforgiving, as an agent would likely not wish to interact further with a partner that profited at his expense. We next claim that consistent strategies are a plausible and interesting subclass of strategies. They are likely to arise in practice because they are simple to implement, and they also give rise to an interesting interpretation of society in which there are good and bad people

[1] This is without loss of generality under the condition that reputation systems (or other forms of information flow) cannot be influenced by using multiple names, since an agent would want to maximize utility for each pseudonym independently.

[2] See Section 4 for a more detailed discussion.

in the world. However, we do not rely on such justifications and interpretations for the restriction of our strategy space. We prove that if agents are trusting, it is an equilibrium for them to be consistent for a wide range of parameters (and in fact, for many ranges of parameters, all equilibria involve only consistent strategies). Finally, we prove that the social norm of trusting strangers is in fact equilibrium behavior for a wide range of parameters (e.g. when the marginal gain from defection is sufficiently small), assuming that agents are consistent. To summarize, we prove that, for a wide range of parameters, it is sufficient to assume either that agents are consistent or that agents are trustworthy. The remaining two assumptions, and hence our characterization of equilibria, then follow as equilibrium behaviour.

In summary, this paper introduces a model of networked Prisoner's Dilemma in a society of anonymous agents, such as agents interacting over the Internet. In this context, social capital is a vehicle to support cooperative behavior, and thus equilibria with non-trivial fractions of cooperators. This effect arises endogenously with no explicit mechanism for reputation tracking or cooperation enforcement. The equilibria we derive are stable in the sense that, if the proportion of cooperators is perturbed, then new entrants have strict preferences for cooperation or defection, bringing the system back to equilibrium.

2. A SOCIAL MODEL OF COOPERATION

The object of study is a prisoner's dilemma interaction governed by the following payoff matrix.

	C	D
C	1,1	-b,1+a
D	1+a,-b	0,0

We take $a, b > 0$ and $a - b < 1$ so that, while mutual cooperation is the uniquely efficient outcome, defection is a dominant strategy. Individuals are matched via a random network formation process and play the stage game iteratively with their partners. We explore the potential to sustain cooperative behavior via the mechanism of thereby attracting a greater number of partners, and hence higher total payoff.

2.1 The model fundamentals

There is a (finite or infinite) countable set of agents interacting with each other in discrete time. Each agent alive at time t survives to $t + 1$ (independently) with probability $\delta \in [0, 1)$. Whenever an agent dies, it is replaced by a newborn agent.

Each agent sponsors k connections to other agents (his *outlinks*), for some $k \geq 1$. The connections of an agent that are sponsered by others are termed his *inlinks*. Each connection persists until one of the partners dies or defects.[3] When a connection is broken, the agent who sponsored it rematches with another agent, chosen uniformly at random, at the next time period.[4]

[3]This assumption is relaxed in Section 4.

[4]The assumption that partners are chosen uniformly at random leads to a very simple network structure. However, our conclusions remain qualitatively unchanged if we extend the network formation model to include additional rules, such as preferential attachment and search via short random walk, that result in graphs that more closely resemble real-world social networks. See Section 6.

At birth, an agent chooses whether to be a cooperator or a defector. The decision is made rationally so as to maximize expected discounted lifetime payoffs from social interactions. Moreover, the decision is taken with commitment: the agent plays the same strategy in each game throughout its life.[5]

To summarize, each time period proceeds according to the following order of events:

1. New agents are born.
2. Each agent attaches its free links to other agents.
3. The game is played and payoffs are realized.
4. Agents sever all links to partners who defected.
5. Death occurs.

2.2 Optimal behavior choices

Given the model as specified above, we now compute expected discounted utilities to the choices of cooperation and defection. These utilities depend on the model's parameters, (a, b, δ), as well as the proportion of cooperative agents in society, q, determined endogenously.

We begin by making a few simple observations. An agent who defects loses all of its connections each period. Thus, its optimization problem is identical at every date, provided the system is at steady-state so that q is not changing. Second, if an agent's partner defects, then given the assumption of strategies with commitment, it is rational to believe the agent will continue to defect forever, and hence to sever the link to him in exchange for a random partner.

The main task in computing utilities is to keep track of the fraction of inlinks and outlinks between agents of different behaviors. We associate an agent's behavior with his type, C or D. Define $n^{Out}_{XY}(s)$ as the expected fraction of outlinks from an agent of type X at age s to agents of type Y, $X, Y \in \{C, D\}$. The fraction of links between cooperators can be computed recursively according to

$$n^{Out}_{CC}(s) = \delta n^{Out}_{CC}(s-1) + q(1 - \delta n^{Out}_{CC}(s-1)).$$

The first term retains the existing links with cooperators who remain alive, while the second term takes all links from the previous period that were broken (due to death or defection) and re-wires them, getting a fraction q of new cooperators. Setting $n^{Out}_{CC}(-1) = 0$ and solving produces

$$n^{Out}_{CC}(s) = q\left(\frac{1 - (\delta(1-q))^{s+1}}{1 - \delta(1-q)}\right).$$

The remaining links sponsored by a cooperator go to defectors, so that $n^{Out}_{CD}(s) = 1 - n^{Out}_{CC}(s)$. For defectors, as mentioned, the case is much simpler, and depends only on the population frequency of cooperators. We have $n^{Out}_{DC}(s) = q$ and $n^{Out}_{DD}(s) = 1 - q$.

We next need to compute the expected fractions of an agent's inlinks from both types of nodes. To do so, we first need to compute the distributions of outgoing links from agents of different ages. First, notice that the probability that a randomly selected node is age s is $p(s) = (1 - \delta)\delta^s$. Then, the total per-agent mass of connections re-wired by agents of each type at any given time are

$$r_C = \sum_{s=0}^{\infty} qp(s)\left(1 - \delta n^{Out}_{CC}(s-1)\right) = \frac{q(1-\delta^2)}{1 - \delta^2(1-q)},$$

$$r_D = \sum_{s=0}^{\infty}(1-q)p(s) = 1 - q.$$

[5]Again, this assumption is relaxed in Section 4.

This can be seen as follows. The fraction of C (D) agents is given by $qp(s)$ ($(1-q)p(s)$). The mass of links re-wired by C agents of age s is $1-\delta n_{CC}^{Out}(s-1)$, and for D agents is 1.

We can now compute the expected fraction of inlinks. Define $n_{XY}^{In}(s)$ as the expected fraction of inlinks an agent of type X at age s has from agents of type Y, $X,Y \in \{C,D\}$. For CC links, we have the recursive relationship

$$n_{CC}^{In}(s) = \delta n_{CC}^{In}(s-1) + r_C.$$

Setting $n_{CC}^{In}(-1) = 0$ and solving produces

$$n_{CC}^{In}(s) = r_C \frac{1-\delta^{s+1}}{1-\delta}.$$

The remaining calculations are straightforward since they all involve defectors whose links are re-set every period. We have $n_{CD}^{In}(s) = n_{DD}^{In}(s) = r_D$ and $n_{DC}^{In}(s) = r_C$.

Finally, we can now define the expected lifetime utility of choosing to be a perpetual cooperator or defector. For convenience we scale these utilities by $1/k$. First we compute the expected payoff at a particular age s by summing the payoffs over his expected set of connections. We have

$$\begin{aligned}\pi_C(s) &= \left(n_{CC}^{Out}(s) + n_{CC}^{In}(s)\right) - b\left(n_{CD}^{Out}(s) + n_{CD}^{In}(s)\right),\\ \pi_D(s) &= (1+a)\cdot\left(n_{DC}^{Out}(s) + n_{DC}^{In}(s)\right).\end{aligned}$$

Expected normalized discounted lifetime utilities are then simply $u_X = (1-\delta)\sum_{s=0}^{\infty}\delta^s \pi_X(s)$, $X \in \{C,D\}$. Simplifying the expressions delivers

$$\begin{aligned}u_C &= \frac{2q - b(1-q)(2-\delta^2(2-q))}{1-\delta^2(1-q)},\\ u_D &= \frac{(1+a)q(2-\delta^2(2-q))}{1-\delta^2(1-q)}.\end{aligned}$$

3. EQUILIBRIUM CHARACTERIZATION

Each agent chooses at birth C or D so as to maximize his expected discounted utility. In order to characterize optimal choices we are interested in comparing u_C and u_D as a function of q under various parameterizations of the model. It will be convenient to define $\Delta(q;a,b,\delta) = u_D - u_C$.

Notice that $q = 0$ is always stable, in the sense that $\Delta(0;a,b,\delta) > 0$. That is, if there are sufficiently few cooperators, then it cannot be optimal to commit to cooperation. On the other hand, if $q = 1$, then we will see that cooperation is sustained in certain settings, e.g., if the expected lifetime is sufficiently large and the gain from defecting against a cooperater is sufficiently small, then the long-term value from accumulated cooperater links is out-weighed by the short-term gain from defection.

We are interested in characterizing the mixtures of cooperate and defect types that can be sustained in a *stable* steady state of the system. A steady state is stable if the system always returns to it after small disturbances. In general, the stable states of the system (generically) fall into three categories. First, it is possible that $\Delta(q;a,b,\delta) > 0$ for all q, in which case all-defection is the unique stable state. For any a,b, this will be the case for sufficiently small δ. Second, it may be the case that there is a unique q^* for which $\Delta(q;a,b,\delta) = 0$, above which $\Delta(q;a,b,\delta) < 0$. In this case, all-defection ($q = 0$) and all-cooperation ($q = 1$) are the two stable states, and q^* is an unstable steady state. Finally, it may be the case that $\Delta(q;a,b,\delta) < 0$ for an interior region

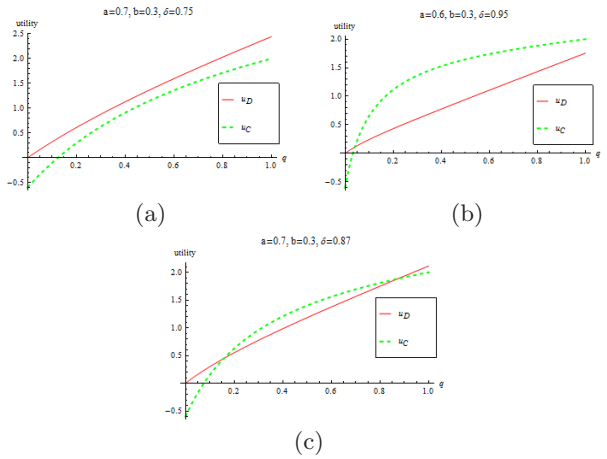

Figure 1: Utility curves corresponding to different patterns of equilibrium occurance. (a) Only the all-defection state is an equilibrium. (b) The all-cooperate and all-defect states are at equilibrium, and there is an unstable equilibrium for some $q \in (0,1)$. (c) The all-defect state is an equilibrium, as well as two interior equilibria: the rightmost stable, the leftmost unstable.

of $q \in (\underline{q}, \overline{q})$, and positive otherwise. In this case $\overline{q}$ is a stable state that involves the co-existence of cooperators and defectors (and $\underline{q}$ is an unstable steady state). See Figure 1 for an illustration of utility curves u_C and u_D corresponding to each of these scenarios. The following result characterizes these possibilities.

PROPOSITION 1. *The all-defection state ($q = 0$) is always an equilibrium. The remaining equilibria are as follows:*
If $a > 1$, then:

(i) if $b < 2$ and δ is sufficiently large, there exist two interior equilibria: one stable and one unstable, with the stable equilibrium occurring with more cooperators.

(ii) otherwise ($b \geq 2$ or δ not large enough) there is only the $q = 0$ equilibrium.

If $a < 1$, then:

(iii) if δ is sufficiently large then the $q = 1$ state is an equilibrium, and there will exist an unstable internal equilibrium.

(iv) if δ is sufficiently small then only the $q = 0$ state is an equilibrium.

(v) if $b < a(1+a)$, then there exists an intermediate range of δ for which there are two interior equilibria: one stable, and one unstable.

PROOF. First, notice $\Delta(0;a,b,\delta) = 2b > 0$, so that $q = 0$ is always an equilibrium.

Second, $\Delta(1;a,b,\delta) = 2a - \delta^2(1+a)$, which is negative whenever $a < \frac{\delta^2}{2-\delta^2}$. Thus, when $a < 1$, $q = 1$ is an equilibrium for large enough δ, in which case there is an interior q at which $u_C = u_D$. When $a > 1$, $q = 1$ is never an equilibrium.

Next, internal equilibria must satisfy the condition $\Delta(q;a,b,\delta)=0$. Solving for δ produces

$$\delta^*(q;a,b)=\sqrt{\frac{2\left(aq+b(1-q)\right)}{(2-q)\left((1+a)q+b(1-q)\right)}}.$$

Clearly, given $a,b>0$, $\delta^*(q;a,b)$ is bounded away from zero for all q. Furthermore, taking derivatives we see that for any $a,b>0$ and $q\in[0,1]$, $\Delta(q;a,b,\delta)$ is strictly decreasing in δ. Hence, for sufficiently small δ ($\delta<\sqrt{\frac{a}{1+a}}$ suffices), u_D dominates u_C for all q, and the only equilibrium is $q=0$.

It is easily seen that $\delta^*(0;a,b)=1$ and $\delta^*(1;a,b)=\sqrt{\frac{2a}{1+a}}$, which is less than one if and only if $a<1$. Also,

$$\begin{aligned}\frac{\partial\delta^*(q;a,b)}{\partial q}|_{q=0} &= \frac{b-2}{4b},\\ \frac{\partial\delta^*(q;a,b)}{\partial q}|_{q=1} &= \frac{a(1+a)-b}{(1+a)^2\sqrt{\frac{2a}{1+a}}}.\end{aligned}$$

As is clear from the representation above, $\delta^*(q;a,b)$ has a unique point of discontinuity, and it is strictly greater than one. Thus, $\delta^*(q;a,b)$ is continuous on the unit interval. It is also continuously differentiable on the same interval. We now show the following

Claim: $\delta^*(q;a,b)$ has at most one local optimum in the interval [0,1].

Notice that because $\delta^*(q;a,b)$ is continuously differentiable, the claim implies that it is either monotonic on [0,1] (in the case of having no optima), or it is "single-peaked" on [0,1] (in the case of one optimum).

Proof of claim: Because $\delta^*(q)$ is bounded away from zero, its derivative has the same zeros as the derivative of $(\delta^*(q))^2$. Setting $\frac{\partial(\delta^*(q))^2}{\partial q}=0$ produces a quadratic in q. Call the solutions q^+ and q^-. We must show that at most one solution falls inside the unit interval.

If $a=b$ then it is easy to see that $q^+=q^-=1-b/2$.

If $a=b+1$ then $q^-=-b-\sqrt{b(b+2)/2}<0$.

Otherwise, $q=\frac{-b}{a-b}\pm\frac{\sqrt{(1+a-b)(2a-b)b}}{(1+a-b)(a-b)}$. Call this $Q_1\pm Q_2$. If Q_2 is not real, we are done, so assume it is. If $a-b>0$ then $Q_1<0$, so at least one of the solutions is less than zero. On the other hand, if $a-b<0$ then $q_1>1$ so at least one of the solutions is greater than one. This proves the claim.

An interior equilibrium occurs when there exist $\bar{\delta}<1$ and $0<\underline{q}<\overline{q}<1$ such that $\delta^*(\underline{q};a,b)=\delta^*(\overline{q};a,b)=\bar{\delta}$. This is so because such a situation guarantees that $\Delta(q;a,b,\delta)=0$ for two different values of q at the same value of δ.

From the expressions above, this is the case, for some δ, provided that $b<2$ and $b<a(1+a)$. First consider $a>1$. The condition for interior equilibrium existence reduces to requiring $b<2$. Such an equilibrium exists for all sufficiently large δ because $\delta^*(1;a,b)>1$. This proves parts (i) and (ii).

Next consider $a<1$. Since $a(1+a)<2$, an interior equilibrium exists for some δ whenever $b<a(1+a)$. However, now it is the case that $\delta^*(1;a,b)<1$, which implies that an interior equilibrium does not exist for $\delta>\delta^*(1;a,b)$. For $\delta>\delta^*(1;a,b)$, there is an unstable interior equilibrium, and $q=1$ is a stable equilibrium. □

Each of the above conditions occurs for reasonable ranges of parameters; see Figure 1 for some typical examples.

Let us now provide some intuition behind the results of Proposition 1. Observe that defectors and cooperators gain utility in very different ways. A defector gains utility directly by interacting with cooperators and exploiting them. Their links do not persist from one turn to the next. Thus the per-period utility of a defector is (practically) proportional to the fraction of cooperators in the system. Parameter a dictates the rate at which utility increases with cooperators.

By contrast, a cooperator gains utility by building a network of relationships from which he can extract utility over his lifetime. Given sufficient time, the neighborhood of a cooperator limits to a critical size, at which point the rate of decay of existing friends matches the rate of finding other cooperators. A major factor in the payoff of a cooperator is the amount of time necessary to approach this critical neighborhood size, relative to the expected lifespan. This quantity is influenced by the number of cooperators in the system, but this influence suffers diminishing returns: when there are few cooperators present, a small increase will have large effects on the number of cooperators expected to meet each other; when there are many cooperators, they will all likely reach their critical neighborhood sizes, and thus the addition of more cooperators has little effect.

The utility of a cooperator will also be affected by the losses he incurs from interacting with defectors. This is a linear effect, proportional to the number of defectors in the system, similar to the total utility gained by a defector. Parameter b determines the rate at which utility decreases with the number of defectors.

When $a>1$, the expected utility of a single defector in an otherwise all-cooperator environment will be greater than the expected utility of a cooperator who has a full neighborhood of other cooperators. That is, a $q=1$ equilibrium cannot exist for any δ. Starting from the $q=1$ state, defectors begin to enter the system. As more defectors enter, the expected utility of each defector decreases linearly. How is the utility of the cooperators affected?

First, if b is very large, the presence of more defectors degrades the utility of the cooperators heavily, due to losses that occur when interacting with defectors. If b is large enough, this degradation will be so severe that defecting will always be the superior strategy, and the only equilibrium of the system will be at $q=0$.

Second, if the expected lifetime is sufficiently small, the presence of more defectors will make it noticeably less likely that cooperators will form full neighborhoods of other cooperators within their lifetimes, again degrading their utility and destroying the $q=1$ equilibrium. If δ is small enough, the payoff due to forming a (partial) neighborhood will never overtake the utility of defecting, and again the only equilibrium of the system will be the $q=0$ state.

Third, if b is small and δ is sufficiently large, then an increase in the number of defectors will have a small effect on the expected welfare of a cooperator. Thus, as more defectors enter the system, the gap in welfare between defectors and cooperators will close, until at some interior point they become equal. This is precisely the stable interior equilibrium described in the first half of the proposition.

When $a<1$, the expected utility of a single defector in an otherwise all-cooperator utopia will be less than the expected utility of a cooperator who has a full neighborhood of other cooperators. That is, an all-cooperate equilibrium exists provided δ is sufficiently large. In such a case, there must also be an unstable internal equilibrium (since both the $q=0$ and $q=1$ states are stable, there must be some

interior state where utilities are equal).

If δ is very small, then cooperators will not expect to find each other during their lifetimes. In such a setting, it will always be better to defect than to cooperate, and only the $q = 0$ state will be stable.

In reference to the last case of the proposition, consider a range of δ for which cooperators are not guaranteed to fully reach their critical neighborhood levels, but will come close. It may then be the case that a defector gains more utility than a cooperator in the $q = 1$ state. However, if the losses incurred due to exploitation are not too large, and if δ is large enough that cooperators expect to find many other cooperators over their lifetimes (though not as many as they could hope for), then an increase in the number of defectors will have more effect on the defectors' utilities than on the cooperators' utilities. In this case, starting from $q = 1$ and adding defectors to the system, one reaches a state where the utilities of the defectors and the cooperators are equal.

In summary, stable interior equilibria occur whenever

(a) defecting is preferable to cooperating in a world where all agents cooperate, and

(b) when defectors enter a mostly-cooperator system the rate of decay of defector utility is greater than the rate of decay of cooperator utility.

Condition (b) generally requires that parameter b not be too large, and that δ not be too small. Condition (a) requires either that parameter a be large, or that δ not be too large.

4. EXTENSIONS

The equilibrium identified in Proposition 1 can be thought of as arising under a very natural social norm, in the spirit of Ghosh and Ray (1996). The social norm specifies how to behave in one's relationships as well as how to manage one's relationships. That is, under the prescription of ending relationships upon observing a defection, the configurations identified in Proposition 1 are optimal, and under the behavior described by the proposition, ending relationships upon observing a defection is rational.

4.1 Maintenance of relationships

We address first the norm that individuals sever a relationship upon observing a defection. Under the strategies we consider, namely life-long cooperation and life-long defection, the beliefs of an individual regarding the future play of his partners are easy to describe. In fact, after a single interaction, the individual can perfectly forecast his partners' future play. Therefore, it is optimal to always maintain a relationship after observing cooperation, and it is optimal to sever a relationship after observing defection. We are precluding the possibility of deviations from these two strategies, so there is no consideration of off-path beliefs. We now turn to the consideration of enriching the strategy space.

4.2 More strategies

The main analysis was conducted under the assumption that individuals have available to them only two strategies at their birth. Optimality, then, requires taking rational expectations over the implied outcomes of these two actions, and choosing appropriately. There is no consideration of deviations; the choice is assumed to be made with commitment. We now want to show that our analysis is robust to allowing for fully general strategies.

First notice that for a defector, the decision problem at every point in time is identical. This is so because, under the norm regarding relationship maintenance, he loses all of his connections at every period. Thus, if he decides today that perpetual defection is better than perpetual cooperation, he will reach the same conclusion tomorrow.

For a cooperator the situation is complicated by the fact that his state (i.e. number of in-links and out-links) changes over time. In expectation, a cooperator is at least as happy with his choice, at birth, than he would be under the alternative plan of defection. But there may arise interim situations in which a cooperator prefers to defect in a particular period, after which his optimization problem is identical again to the one at his birth.

We now introduce notation to describe the more general set of behavioral strategies we have in mind, which are mappings from a node's history into a current choice. The outcome from any single relationship in a given time period is $h_t^r \in \{C, D\}^2$. The fact that a defection by either partner ends a relationship simplifies the set of strategies that can be realized. In particular, every current relationship must have a history of exclusively (C, C) outcomes. Let $h_t^O = \{h_t^k\}_{k \in K_O}$ denote the outcome of each out-link relationship at time t. Let K_t^I denote the mass of in-links at time t, and let $h_t^I = \{h_t^k\}_{k \in K_t^I}$ denote the outcome of each in-link relationship at time t. Then the outcome at time t is $h_t = h_t^O \cup h_t^I$, and the node's history at time s is $h_s = (h_0, h_1, \ldots, h_t)$. Denote the space of histories at time s by H_s and the space of all histories by $H = \cup_{s<\infty} H_s$. A strategy for a node specifies how to play at every history. Define $\phi : H \to [0, 1]$, with the interpretation that $\phi(h)$ dictates the probability to play C at history h.

The next result provides a sufficient condition to guarantee that cooperators never have a profitable deviation.

PROPOSITION 2. *Assume that*

$$\frac{1+b}{1-(1-q)\delta^2} \geq 1 + a. \tag{1}$$

Then the strategies from Proposition 1 constitute equilibria under fully general strategies given the norm of severing relationships with defectors.

PROOF. We prove that Condition 1 guarantees that a cooperator gains more than a defector from the accumulation of (in- or out-) links with cooperators. Thus, at any point in its life, a node that found it optimal to cooperate at birth finds it strictly optimal to cooperate later on.

Take a node who has existing links with cooperators. Define k_I and k_O to be the mass of in- and out-links, respectively, that the node has to cooperators. We want to show that the utility gain to a cooperator from being given k_I and k_O is greater than the utility gain to a defector, relative to the situation at birth, when $k_I = k_O = 0$.

The utility gains to a node are additively separable in k_I and k_O, and so we analyze them separately. The utility gain to a defector from k_I in-links with cooperators is $\Delta u_D(k_I) = (1+a)k_I$, since the defector loses these links after his first defection. For cooperators, the gain is $\Delta u_C(k_I) = \frac{k_I}{1-\delta^2}$. This is so because the cooperator gets extra utility for each period of the life of the relationship. We have that $\Delta u_C(k_I) > \Delta u_D(k_I)$ whenever $\frac{1}{1-\delta^2} > 1 + a$, which is necessary to sustain cooperation in equilibrium anyway.

We turn now to out-links, where a fraction k_O of the

nodes out-links are matched to cooperators, and the remaining out-links are matched to the population at random. For defectors, $\Delta u_D(k_O) = (1+a)(1-q)k_O$. To see this, note that the gain in out-links to cooperators is $k_O+(1-k_O)q-q = (1-q)k_O$, and this gain is realized for exactly one period. For cooperators, $\Delta u_C(k_O) = \frac{(1+b)(1-q)k_O}{1-(1-q)\delta^2}$. To see this, notice that per interaction, a cooperator gains $1+b$ from interacting with a cooperator rather than a defector, and the node gains $k_O + (1-k_O)q$ extra out-links to cooperators. Finally, each of these relationships survives the period with probability δ^2 and when it ends is replaced by a relationship with a defector with probability $1-q$. Setting $\Delta u_C(k_O) > \Delta u_D(k_O)$ completes the proof. □

The condition in Proposition 2 guarantees that, as a node obtains more in-links and out-links with cooperators, the gain from those relationships is higher to a cooperator than to a defector. Thus, a node that found it optimal to cooperate at birth necessarily finds it optimal to cooperate at any point in its lifetime. Thus, under this condition, the situations outlined in our main result constitute equilibria of the repeated game under fully general strategies.

Even when the condition is violated, the possibility for profitable deviations come in a very limited form. First, notice that the condition is always satisfied when $b \geq a$. Thus, the only possibility for profitable deviation occurs when $a - 1 < b < a$. In that case, we require that δ be sufficiently large in order to rule out profitable deviations. Next, notice that incentive to defect is strongest when the number of out-links to cooperators is high relative to the number of in-links from cooperators. This is so because, in order to want to defect, there must be some cooperators to cheat. It is better to defect when those cooperators come from out-links, since those are the ones that are easier to replace over the remaining lifetime.

Recall that there is no scope for a profitable deviation under our condition. When the condition is violated, a cooperator has a profitable deviation when the ratio of his out-links with cooperators to his in-links from cooperators is sufficiently high. Define this ratio to be $K = k_O/k_I$. We stress that under the expected conditions, such situations do not arise. In a model with a continuum of agents and links, the expected frequencies obtain almost surely, and the results from our main proposition hold. However, when agents maintain only a finite number of links, cooperators will reach a state that gives them a profitable deviation with positive probability. This happens to a node, for instance, whenever all the cooperators maintaining links to him die simultaneously. In practice, these situations have significant probability only very early in the life of a cooperator, before it has had time to build a network of in-links.

PROPOSITION 3. *Assume that* $\frac{1+b}{1-(1-q)\delta^2} < 1+a$. *Then a cooperator has a profitable deviation if and only if*

$$K\left[(1+a) - \frac{1+b}{1-(1-q)\delta^2}\right] > \frac{1}{1-\delta^2} - (1+a).$$

PROOF. If the right hand side is negative, then defection dominates cooperation, so assume otherwise. Then, the right hand side is the extra gain that a cooperator realizes from an in-link with a cooperator relative to the gain a defector realizes. The term in brackets is the extra gain a defector realizes from an out-link to a cooperator relative to the gain a cooperator realizes, which is positive by assumption. Then the result simply expresses that when the ratio of out-links to in-links is high enough, the net gain to defection is positive. □

4.3 Accepting links

We next address the norm that each agent, whether cooperator or defector, will choose to accept any relationship initiated by another. This is certainly reasonable behaviour from a defector, since defectors always obtain non-negative utility from any relationship. For cooperators, however, the rationality of this norm is far less clear. One might imagine a scenario in which there are many defectors in the population, and moreover a newly proposed link is likely to have come from a defecting agent. In such a case, it may be that a cooperator suffers an expected utility loss from accepting an incoming link, and hence a rational agent should refuse all relationship invitations. Of course, such decisions have a severe impact on the network, as they prevent the formation of any profitable relationships. With this in mind, we wish to characterize the circumstances in which such a scenario can occur at a stable equilibrium in our model.

Recall from Proposition 1 that we may have equilibria at $q = 0$ and possibly $q = 1$, depending on our choice of model parameters. However, the question of whether or not to accept incoming links is not interesting in these cases, as either there are no cooperators to deviate from the norm (when $q = 0$), or the utility of accepting inlinks is trivially positive (when $q = 1$). We therefore limit our attention to internal stable equilibria.

Our first result is that, at any internal stable equilibrium for which q is at least $2/3$, each cooperator has positive expected utility from accepting an incoming link. That is, if at least $2/3$ of the population is cooperating, then without loss of generality one can assume that rational agents are trusting and will enter into relationships proposed by others.

PROPOSITION 4. *Suppose that* $q \in (0,1)$ *is an interior stable equilibrium. If* $q > \frac{2}{3}$, *a cooperator obtains positive expected utility from accepting an in-link.*

PROOF. Fix a and b, and choose δ such that there exists a stable equilibrium $0 < q < 1$. The condition that rationalizes accepting in-links is that $r_C - b * r_D \geq 0$, which is equivalent to $b \leq \frac{q}{(1-q)(1-(1-q)\delta^2)}$.

If $q > \frac{2}{3}$, then $\frac{q}{(1-q)(1-(1-q)\delta^2)} \geq \frac{q}{1-q} \geq 2$. Furthermore, since we assume that a stable interior equilibrium exists, Proposition 1 implies that $b \leq 2$ (either directly, if $a > 1$, or from the fact that $b < a(1+a) < 2$ if $a < 1$). Thus $b \leq \frac{q}{(1-q)(1-(1-q)\delta^2)}$ as required. □

Our next result is that, if $a < 1$, then *any* stable equilibrium satisfies the property that cooperators maximize their expected utility by accepting incoming links. In other words, if the relative gain of a defector is not too large, then rational cooperators will choose to be trusting at equilibrium, regardless of the number of defectors in the population.

PROPOSITION 5. *Suppose that* $a < 1$ *and that* $q \in (0,1)$ *is a stable interior equilibrium. Then a cooperator obtains positive expected utility from accepting an in-link.*

PROOF. Fix a and b and choose δ such that an internal stable equilibrium q exists. As in Proposition 4, it suffices to show that $b \leq \frac{q}{(1-q)(1-(1-q)\delta^2)}$.

Since $a < 1$, Proposition 1 implies that $b < a(1+a)$, and hence $b < 2a$ and $b < 1+a$. Recall from the proof of Proposition 1 that at a stable equilibrium we have

$$\delta^2 = \frac{2(aq + b(1-q))}{(2-q)((1+a)q + b(1-q))}. \quad (2)$$

Define function $Z(q; a, b)$ by

$$Z(q; a, b) := \frac{(2-q)((1+a)q + b(1-q))}{(1-q)(2-q+aq+b(1-q))}.$$

Substituting (2), it can be verified that $\frac{q}{(1-q)(1-(1-q)\delta^2)} = Z(q; a, b)$. It therefore suffices to show that $b \leq Z(q; a, b)$ at all stable internal equilibria.

We first claim that $Z(q; a, b)$ is monotonic non-decreasing as a function of q. This follows immediately from the following expression for the derivative of Z with respect to q,

$$\frac{\partial Z(q; a, b)}{\partial q} = \frac{q}{(1-q)^2} + \frac{4a - 2b}{(2-q+aq+b(1-q))^2},$$

which is non-negative since $b < 2a$. Let q^*_{min} denote the minimal q at which a stable equilibrium occurs, over all possible choices of δ. Since $Z(q; a, b)$ is non-decreasing in q, it is sufficient to show that $b \leq Z(a, b, q^*_{min})$.

We next derive an expression for q^*_{min}. Recall from the proof of Proposition 1 that function $\delta^*(q; a, b)$, which relates δ to q at equilibrium, is concave and single-peaked in the range $(0, 1)$. Furthermore, whenever there exist $0 < \underline{q} < \overline{q} < 1$ such that $\delta = \delta^*(\underline{q}; a, b) = \delta^*(\overline{q}; a, b)$, $\overline{q}$ is a stable equilibrium and $\underline{q}$ is not. We conclude that q^*_{min} is precisely the value of q at which function $\delta^*(q; a, b)$ achieves its minimum on $[0, 1]$. Solving $\frac{\partial \delta^*(q;a,b)}{\partial q} = 0$ for q, we obtain the pair of solutions

$$q = \frac{-b \pm \sqrt{\frac{b(2a-b)}{1+a-b}}}{a-b}.$$

Write $r(a, b) := \sqrt{\frac{b(2a-b)}{1+a-b}}$. Using the facts that $a < 1$ and $b < a(1+a)$, it is a simple exercise to show that $\frac{-b+r(a,b)}{a-b} \in [0, 1]$ and $\frac{-b-r(a,b)}{a-b} \notin [0, 1]$. We conclude that

$$q^*_{min} = \frac{-b + r(a, b)}{a - b}.$$

Substition and simplification then yields

$$Z(q^*_{min}; a, b) = \frac{b(2(1+a) - b)^2}{L(a, b)}$$

where

$$\begin{aligned} L(a, b) = {} & 2(1+a)^2 r(a, b) \\ & - (1+a)b(2(1-a) + r(a, b)) \\ & + b^2(r(a, b) - (1+a)) \end{aligned}$$

Thus, to show that $b < Z(q^*_{min}; a, b)$, it suffices to show

$$\begin{aligned} (2(1+a) - b)^2 > {} & 4(1+a)^2 \left(\frac{r(a, b)}{2}\right) \\ & - 4(1+a)b\left(\frac{2(1-a) + r(a, b)}{4}\right) \\ & + b^2(r(a, b) - (1+a)). \end{aligned} \quad (3)$$

We will derive (3) with the help of the following claim:

Claim: For all $a < 1$ and $b < a(1+a)$, it must be that $r(a, b) < 2$ and $r(a, b) < 1 + a$.

To prove the claim, we note that for fixed a, $r(a, b)$ attains its maximum at $b = 1 + a \pm \sqrt{1-a^2}$. Since $b < 1 + a$, the admissible solution is $b = 1 + a - \sqrt{1-a^2}$, which yields $r(a, b) = \sqrt{2 - 2\sqrt{1-a^2}} < 2\sqrt{a}$. But $2\sqrt{a} < 2$ since $a < 1$, and moreover $2\sqrt{a} < 1 + a$ by considering the fact that $(1 - \sqrt{a})^2 \geq 0$. Thus $r(a, b) < 2$ and $r(a, b) < 1 + a$ as required, completing the proof of the claim.

Our claim immediately implies that $\frac{r(a,b)}{2} < 1$, $r(a, b) - (1+a) < 1$, and $\frac{2(1-a)+r(a,b)}{4} \in (0, 1)$. Taking

$$\lambda = \max\left\{\frac{r(a, b)}{2}, r(a, b) - (1+a), \frac{2(1-a) + r(a, b)}{4}\right\},$$

we conclude that

$$\begin{aligned} & 4(1+a)^2\left(\frac{r(a, b)}{2}\right) \\ & \quad - 4(1+a)b\left(\frac{2(1-a) + r(a, b)}{4}\right) \\ & \quad + b^2(r(a, b) - (1+a)) \\ & \leq 4(1+a)^2\lambda - 4(1+a)b\lambda + b^2\lambda \\ & = \lambda(2(1+a) - b)^2 \\ & < (2(1+a) - b)^2 \end{aligned}$$

which is (3), completing the proof of Proposition 5. □

Finally, we note that Proposition 5 fails to hold when we remove the assumption that $a < 1$. Indeed, for any given $a > 1$, there exists a stable equilibrium at which rational cooperators would choose to reject in-links. This follows from the observation that, when $a > 1$, a stable equilibrium exists for any $b < 2$ and $\delta < 1$; however, as $b \to 2$ and $\delta \to 1$, the value of q for this equilibrium becomes arbitrarily small. The quantity $\frac{q}{(1-q)(1-(1-q)\delta^2)}$ from Proposition 4 can then be made arbitrarily close to 1, and hence less than b. For example, if we choose $a = 1.9$, $b = 1.6$, and $\delta = 1 - \frac{1}{1000}$, then a stable equilibrium occurs at $q < 0.3$, from which it can be verified that $r_C - b * r_D < 0$, meaning that the expected utility from accepting an in-link is strictly negative.

In summary, the norm that cooperators accept all in-links is without loss for rational agents at a stable equilibrium whenever there are sufficiently many cooperators in the network. If $a < 1$, it turns out that *any* stable equilibrium must have enough cooperators to motivate that acceptance of in-links. For the case of $a > 1$, there exist stable equilibria with arbitrarily few cooperators, and hence there are choices of parameters for which rational agents would choose not to accept incoming links.

5. RELATED LITERATURE

There is a large body of work seeking to explain the prevalence of cooperation in social networks. One strain of research models the emergence of cooperation through endogenous affects of modeling assumptions whereas the other looks explicitly at mechanisms, like those of eBay, introduced to enforce cooperation. We discuss each in turn.

5.1 Emergence of Cooperation

The literature on the emergence of cooperation in social networks centers around the Prisoner's Dilemma, the bi-

matrix game described in Section 2. In this game cooperation is the pareto-efficient outcome, and yet defection is a dominant strategy. Game theory predicts that agents will defect in this game, though various experiments and simulations, as well as psychological intuition, indicate that agents tend to cooperate at least partially (a few representative works include [5, 28, 30]).

Much work has gone into exploring theoretical models that explain the prevalance of cooperation in Prisoner's Dilemma games. We briefly outline a few approaches here including repetition with enforcement and the use of "trust".

When the participants of the game are fixed known individuals and the game is repeated an infinite or uncertain number of times, the Folk Theorem of the repeated game literature shows that any mutually beneficial payoff structure can be sustained in equilibrium through *personal enforcement* with the use of threats [1, 4, 14, 15, 17, 33]. For example, players can agree to cooperate with the understanding that, should either defect, the opponent will punish the deviator by defecting forevermore. A variety of extensions discuss how to sustain cooperation with n-player games, finite-horizon games with incomplete information, and games with imperfect observations [17, 16, 27], all for settings in which the opponents are fixed and known throughout time.

When Prisoner's Dilemma is played in a large population, people tend to play games with different opponents over time and personal enforcement threats become impossible to enforce. In fact, players may not even know each others' names. Nonetheless, full cooperation can still be sustained through the use of *community enforcement* [12, 19, 25, 29]. Here, if an agent defects against an opponent, then the opponent starts defecting, initiating a cascade of defect actions and giving rise to a defect contagion which eventually punishes the original deviator. The threat of this defect contagion sustains cooperation in the society as a whole. When anonymity is due to the use of pseudonyms that can be changed, an alternative approach to community enforcement is to penalize all new players for a short number of rounds. This builds a level of trust in agents by forcing them to "pay their dues" [13]. This penalty de-incentivizes agents from taking new pseudonyms, at the cost of reduced efficiency in interactions with new participants.

When agents have choice in their partners or in the length of the relationship, new insights arise [9, 18, 26, 34]. In developing long-term relationships, agents have the opportunity to gradually build trust with their partners. This trust becomes an asset, and the threat of losing it causes agents to cooperate. This cooperation is achieved without information flow (i.e., no public reputation), and without relying on contagious defect strategies which require small populations and are sensitive to occasional byzantine defection.

In contrast to the above work, our mechanism for sustaining cooperation can be identified with *social capital* (taken to mean an agent's network of partners, see [35] for a survey of various definitions of social capital and their implications). Here, the reason to cooperate comes from the fact that, through cooperation, one can gradually build up a social network of other cooperators. Ours is one of the first models that proposes social capital as a justification for cooperation in Prisoner's Dilemma games.[6] Also, ours is the first fully theoretical model that justifies co-existence of cooperation and defection strategies.[7]

5.2 Reputation Mechanisms

In practice, especially in online social network settings, cooperation is often aided by the introduction of explicit reputation mechanisms (see [10] for a survey or [31] for a short exposition). These mechanisms come in several flavors (see [] for a discussion of some of these mechanisms and related empirical studies). Many of those deployed in practice, like that of eBay, use a simple binary or ternary rating scheme. Such schemes have been shown to be highly effective experimentally and theoretically despite their tremendous simplicity [32, 11]. Nonetheless, they suffer from a variety of drawbacks, including barrier-to-entry, issues with illiciting feedback (especially negative feedback), and difficulties ensuring honest reports [31]. Some of these issues can be circumvented through the use of clever mechanism design. Several authors consider inducing incentive-compatible reputation mechanisms by introducing payments or considering repeated interactions [20, 21, 22, 23]. Other mechanisms have been proposed as well, based on the structure of interactions or decentralized gossip protocols [6, 24]. The design of such robust recommendation systems is impeded in settings such as ours where agents can create multiple pseudonyms (i.e. sybils), though the use of asymmetry in ranking functions can help to offset this issue [8].

An alternate approach to studying reputation systems is to define a set of desired outcomes and then derive the mechanisms that satisfy these. This so-called axiomatic approach, borrowed from the social choice literature, has been used to successfully explain the prevelance of certain reputation mechanisms like PageRank and trust-based recommendation systems [2, 3, 36].

6. CONCLUSION

In this extended abstract we developed a model for interaction in a dynamic anonymous network We found that for many parameter settings, if we assume that agents are trusting then it follows that agents are consistent at equilibrium, and vice-versa. Furthermore, there are stable equilibria supporting either full cooperation or some combination of cooperation and defection. Our model predicts that, for some ranges of parameters, the presence of non-cooperative behaviour in an anonymous system is unavoidable. The existence of non-cooperative agents in the system causes cooperative players to value their social capital, which in turn keeps them from deviating. This is in contrast to other models in which only the all-cooperate equilibrium is stable.

Our model of network formation leads to a simple network structure, as partnerships are chosen uniformly at random. We note, however, that our conclusions are qualitatively unchanged if we extend the network formation model to include variations, such as preferential attachment. In this way, our model can be made compatible with social networks that support properties such as low diameter and exponential degree distribution. A more complete exploration of these settings is left as an open problem.

[6] See [] and included references for notable exceptions; here social capital refers to the size of an agent's "club" and comes at a cost that grows as the social capital grows.

[7] A related work [7] describes a model with partial analytical results and simulation results in which cooperation and defection co-exist. Their model differs from ours in that agents are boundedly rational, and their model gives rise to a continuum of equilibria.

Another direction of future study is to explore the range of agent behaviour when Proposition 2 does not hold, and otherwise cooperative players would allow themselves to deviate under certain circumstances. There may be equilibria in which such "rob the bank" strategies interact in interesting ways. More generally, we feel that further research into the interplay between agent interaction and network formation will lead to a better understanding of the overall dynamics of social networks.

7. REFERENCES

[1] D. Abreu. On the theory of infinitely repeated games with discounting. *Econometrica*, 56(2):383–396, March 1988.

[2] A. Altman and M. Tennenholtz. Ranking systems: the pagerank axioms. In *ACM Conference on Electronic Commerce*, pages 1–8, 2005.

[3] R. Andersen, C. Borgs, J. Chayes, U. Feige, A. Flaxman, A. Kalai, V. Mirrokni, and M. Tennenholtz. Trust-based recommendation systems: an axiomatic approach. In *International World Wide Web Conference (WWW)*, 2008.

[4] R. Aumann and L. Shapley. Long term competition: a game theoretic analysis. mimeo, Hebrew University, 1976.

[5] R. Axelrod. *The Evolution of Cooperation*. Basic Books, New York, 1984.

[6] Y. Bachrach, A. Parnes, A.D. Procaccia, and J.S. Rosenschein. Gossip-based aggregation of trust in decentralized reputation systems. *Automomous Agents and Multi-Agent Systems*, 19(2):153–172, October 2009.

[7] E. Bilancini and L. Boncinelli. The co-evolution of cooperation and defection under local interaction and endogenous network formation. forthcoming in Journal of Economic Behavior and Organization, 2010.

[8] A. Cheng and E. Friedman. Sybilproof reputation mechanisms. pages 128–132, 2005.

[9] S. Datta. Building trust. Mimeo, Indian Statistical Institute, 1993.

[10] C. Dellarocas. The digitization of word-of-mouth: Promise and challenges of online reputation systems. *Management Science*, 49(10):1407–1424, October 2003.

[11] C. Dellarocas. Reputation mechanism design in online trading environments with pure moral hazard. *Information Systems Research*, 16(2):209–230, June 2005.

[12] G. Ellison. Cooperation in the prisoner's dilemma with anonymous random matching. *The Review of Economic Studies*, 61(3):567–588, July 1994.

[13] E. Friedman and P. Resnick. The social cost of cheap pseudonyms. *Journal of Economics and Management Strategy*, 10(2):173–199, 2001.

[14] J. Friedman. A noncooperative equilibrium for supergames. *Review of Economic Studies*, 38:1–12, 1971.

[15] J. Friedman. *Oligopoly and the Theory of Games*. North-Holland, Amsterdam, 1977.

[16] D. Fudenberg, D. Levin, and E. Maskin. The folk theorem with imperfect public information. *Econometrica*, 62(5):997–1039, September 1994.

[17] D. Fudenberg and E. Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554, May 1986.

[18] P. Ghosh and D. Ray. Cooperation in community interaction without information flows. *The Review of Economic Studies*, 63(3):491–519, July 1996.

[19] J. Harrington. Cooperation in one-shot prisoners' dilemma. Mimeo, Johns Hopkins, 1991.

[20] R. Jurca and B. Faltings. An incentive compatible reputation mechanism. In *International Conference on Autonomous Agents and Multiagent Systems*, pages 285–292, 2003.

[21] R. Jurca and B. Faltings. Minimum payments that reward honest reputation feedback. In *ACM Conference on Electronic Commerce*, pages 190–199, 2006.

[22] R. Jurca and B. Faltings. Collusion resistant, incentive compatible feedback payments. In *ACM Conference on Electronic Commerce*, pages 200–209, 2007.

[23] R. Jurca and B. Faltings. Obtaining reliable feedback for sanctioning reputation mechanisms. *Journal of Artificial Intelligence Research (JAIR)*, 29:391–419, 2007.

[24] S.D. Kamvar, M.T. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *International World Wide Web Conference (WWW)*, pages 640–651, 2003.

[25] M. Kandori. Social norms and community enforcement. *Review of Economic Studies*, 59:63–80, 1992.

[26] R. Kranton. The formation of cooperative relationships. *Journal of Law, Economics, and Organization*, 12(1):214–233, April 1996.

[27] D. Kreps and R. Wilson. Sequential equilibria. *Econometrica*, 50:863–894, 1982.

[28] L.B. Lave. An empirical approach to the prisoners' dilemma game. *The Quarterly Journal of Economics*, 76(3):424–436, August 1962.

[29] M. Okuno-Fujiwara and A. Postlewaite. Social norms and random matching games. *Games and Economic Behavior*, 9:79–109, 1995.

[30] A. Rapoport and A.M. Chammah. *Prisoner's Dilemma*. University of Michigan Press, 1965.

[31] P. Resnick, R. Zechhauser, E. Friedman, and K. Kuwabara. Reptutation systems. *Communications of the ACM*, 23(12):45–48, December 2000.

[32] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood. The value of reputation on ebay: A controlled experiment. *Experimental Economics*, 9(2):79–101, June 2006.

[33] A. Rubenstein. Equilibrium in supergames with the overtaking criterion. *Journal of Economic Theory*, 21:1–9, 1979.

[34] J. Sobel. A theory of credibility. *Review of Economic Studies*, 52:557–573, 1985.

[35] J. Sobel. Can we trust social capital? *Journal of Economic Literature*, XL:139–154, March 2002.

[36] M. Tennenholtz. Reputation systems: An axiomatic approach. In *Conference on Uncertainity in Artificial Intelligence (UAI)*, 2004.