Collective Learning in Games through Social Networks

Sanne Kosterman and **Nina Gierasimczuk**

Institute for Logic, Language and Computation
University of Amsterdam, The Netherlands
nina.gierasimczuk@gmail.com

Abstract

This paper argues that combining social networks communication and games can positively influence the learning behavior of players. We propose a computational model that combines features of social network learning (communication) and game-based learning (strategy reinforcement). The focus is on cooperative games, in which a coalition of players tries to achieve a common goal. We show that enriching cooperative games with social networks communication can influence learning towards the social optimum under specific conditions on the network structure and on the existing expertise in the coalition.

1 Introduction

With the rise of the Internet and digital games, communication and education have changed rapidly. Two digital techniques introduced towards active and social learning environments are *serious games* and *online social networks*. Both of them seem to be auspicious methods that stimulate learning, but are thus far treated as two distinct approaches.

Several attempts have been made to computationally model the learning behavior of artificial agents, both in games and in social networks. Information transmission and opinion formation in social networks has already been extensively studied (see, e.g., [Bala and Goyal, 1998; DeGroot, 1974; Easly and Kleinberg, 2010; Jackson, 2008]). In those frameworks agents can acquire new knowledge and adjust their opinions by learning from the knowledge and beliefs of neighbors in a network. The theory of learning in games has been extensively studied by [Camerer and Ho, 1999], [Fudenberg and Levine, 1998], and [Laslier *et al.*, 2001], who all discuss normative paradigms for learning towards an equilibrium in repeated games. [Bush and Mosteller, 1955] and [Erev and Roth, 1995] provide models for learning in stochastic games, by making use of *reinforcement learning*. This line of research has proved useful not only for the study of artificial agents, but also for the understanding of human learning behavior [Erev and Roth, 2014;

Niv, 2009]. Yet the above-mentioned learning paradigms mainly focus on competitive agents, and treat games and networks separately.

In this paper, we study the question of how interaction in a social network between players of a cooperative game can possibly influence their learning behavior. We thereby assume players to act as one grand coalition, trying to maximize the group utility, i.e., we take players to be *group-rational*. Before performing an action in the game, players can communicate with each other in a social network about the estimated quality of joint strategies. Although individuals can only communicate directly with their neighbors, they aim at cooperation with the entire social network society.

We start this paper with a classical graph-theoretical model for learning in social networks in Section 2. Thereafter, in Section 3 we provide a probabilistic aggregation method to model the amalgamation of individual opinions after network communication. In Section 4 we describe the process of learning in stochastic games, relying on a reinforcement method. We combine the three formal frameworks in the novel learning paradigm proposed in Section 5. In Section 6 we discuss the hidden assumptions of this model and we argue that it contributes to future computer simulations as well as psychological experiments on human learning.

2 Social Network Learning

The so-called Lehrer-Wagner model for weighted preference aggregation [Lehrer and Wagner, 1981] can be utilized for modeling opinion forming in a weighted social network. Here, each agent holds an individual belief about *multiple* alternatives. He can update this belief through communication, taking into account the opinion of his neighbors and a *degree of trust* towards his neighbors' expertise.[1]

Formally, let $\mathcal{G} = (N, S, u)$ be the game with $N = \{1, \ldots, n\}$ players, where $S = \{s(1), \ldots, s(k)\}$ is the set of k *joint* strategies, and $u = (u_1, \ldots, u_n)$ is a tuple of individual utility functions $u_i : S \rightarrow [0, 1]$ for each $i \in N$. Before the game starts, each player i holds a probability distribution over the set of joint strategies, $b_i : S \rightarrow [0, 1]$. One

[1]In fact, the present model is an extension of a classical social network model [DeGroot, 1974] for communication about *single* events only.

could think of these probabilities as *subjective degrees of belief*[2], with respect to the *optimality* of some joint strategy. The higher the probability $b_i(s)$ for a certain strategy profile $s \in S$, the more player i considers it likely that the joint strategy s is a *social optimum*, meaning that it entails a maximal sum of individual payoffs.

All private opinions can be reflected in a stochastic $n \times k$-matrix B, that holds the subjective probability values $b_{ij} = b_i(s(j))$ on its entries. Each i-th row $b_{i\bullet}$ thus represents the probability distribution of agent i over the set S. We write B^1 to denote the *initial belief matrix*. Additionally, each agent i assigns a weight $w_{im} \in [0,1]$ to all other group members $m \in N$. The weights can be represented by the stochastic $n \times n$-matrix W. The corresponding social network is thus a *weighted directed graph* $G = (N, E_W)$ where N is the set of nodes (agents), E_W is the set of weighted directed edges, and w_{im} is thus the weight on the directed edge from i to m. We assume the graph to allow for *loops*, which represents agents' self-confidence.

Each agent can now communicate with neighbors in the network and update his individual belief by taking a weighted arithmetic mean of the beliefs of all agents he trusts. When iterating this process, the updated belief of agent i about strategy $s(j)$ after t rounds of communication is given by $b_{ij}^{t+1} = \sum_{m \in N} w_{im} b_{mj}^t$. On a societal level, belief updates after t rounds of network communication are the result of applying the trust matrix W to the belief matrix B^t, i.e., $B^{t+1} = W \cdot B^t$ (or, equivalently, applying the weight matrix t times to the initial beliefs B^1, i.e. $B^{t+1} = W^t \cdot B^1$). Each round of belief updating can thus be described by means of the following algorithm.

Algorithm 1 Network Communication at round t

Input: Weight matrix W; belief matrix B^t
1: for all $i \in N, s(j) \in S$: $b_{ij}^{t+1} := \sum_{m \in N} w_{im} b_{mj}^t$
2: $B^{t+1} := \left(b_{ij}^{t+1}\right)_{n \times k} = W \cdot B^t$
Output: Belief matrix B^{t+1}

---

3 Collective Decision-Making

After network communication and belief updating, the coalition needs to decide which joint strategy to adopt in the game. We introduce a *probabilistic social choice function* (PSCF), that maps individual probability distributions over a set of alternatives to a societal probability distribution. Players in a cooperative game can make use of such a PSCF to aggregate all individual preferences, in order to determine which joint strategy is most optimal according to society. As an illustration, consider a football team briefing before the match starts, while the players collectively decide on a team strategy.

A first definition for probabilistic social choice functions was introduced by [Gibbard, 1977], in which a *preference profile*, i.e., a tuple of individual preference orders, is mapped to a *lottery*, i.e., a single probability distribution over the set of alternatives. Gibbard referred to such functions as *social*

---

[2]Not to be confused with the quantitative notion of belief used in quantitative approaches to Belief Revision Theory.

---

decision schemes (SDSs). We will introduce a variant of this notion, that takes as input the stochastic $n \times k$-matrix B, as introduced in the previous section. We write $\mathcal{B}(n, k)$ for the set of all such stochastic $n \times k$-matrices. The output of a probabilistic social choice function is given by a k-ary row vector $\vec{b}$ that represents a single societal probability distribution over the set of k alternatives. We write $\mathcal{B}(k)$ for the set of such stochastic k-ary row vectors. Formally, a *probabilistic social choice function* is thus a function $F : \mathcal{B}(n, k) \to \mathcal{B}(k)$.

It is worth noting that the PSCF provides a way of dealing with the *Condorcet paradox* (named after Marie Jean Antoine Nicolas de Caritat, le Marquis de Condorcet, 1743–1794), since it will always select a winning candidate on the basis of probabilities. Even if society likes all alternatives equally good (represented by equal probabilities for all candidates), a winning alternative will be chosen at random.

We will discuss two probabilistic social choice functions that differ with respect to the weights that individuals receive. Intuitively, the higher the weight, the more influence an agent can exert on the construction of the societal probability distribution. In a *weighted* PSCF, different individuals can receive different weights; in an *averaged* PSCF all individuals receive equal weights.

3.1 Weighted Preference Aggregation

In [Lehrer and Wagner, 1981] it has been shown that a special kind of a *social welfare function*, the so-called "Allocation Amalgamation Method" (AAM), can be used as a weighted preference aggregation method to provide a societal ordering over a set of k alternatives. The main advantage of such a weighted method for preference aggregation is that it can take into account the expertise of specific group members. We will use this method for constructing a *weighted probabilistic social choice function* (wPSCF), that outputs a societal probability distribution rather than a societal ordering over the set of alternatives.

The determination of the weights assigned to individuals, relies on the assumption that all agents agree on how much weight each group member should receive. In terms of the weight matrix W, this agreement on weights corresponds to every row of the matrix being the same. It therefore suffices to represent the weights in a stochastic row vector $\vec{w} = (w_1, \ldots, w_n)$, in which $w_i \in [0, 1]$ represents the weight that agent i receives from society. A *weighted probabilistic social choice function* (wPSCF) is then a PSCF $F : \mathcal{B}(n, k) \to \mathcal{B}(k)$ given by $F(B) = \vec{w}B = \vec{b} = (b_1, \ldots, b_k)$ so that each $b_j = \sum_{i \in N} w_i b_{ij}$. In words, a wPSCF is thus a mapping from the individual probability values to a weighted arithmetic mean of these values, for each alternative $s(j) \in S$.

One can easily check that a wPSCF satisfies several axiomatic properties from social choice theory, among which *independence of alternatives* (IA), *unanimity* (U), *neutrality* (N), and *anonymity* (A).

In fact, one can check that wPSCFs even satisfy a stronger notion of unanimity called *Pareto optimality* (P), which guarantees that strict unanimous agreement among all individuals about the order of alternatives is reflected in the societal preference. wPSCFs also satisfy *social rationality* (SR) which is

a highly desired property of cooperative games with group-rational players and a common goal [Arrow, 1951].

3.2 Averaged Preference Aggregation

Before [Lehrer and Wagner, 1981] introduced their weighted method for allocation amalgamation, [Intriligator, 1973] and [Nitzan, 1975] proposed a different probabilistic aggregation procedure which they call "the average rule". The average rule can also be seen as a social welfare function that, on the input of the individual probability distributions, outputs a societal ordering over the set of alternatives. We will use this method for constructing an *averaged probabilistic social choice function* (aPSCF), that outputs a societal probability distribution rather than a ordering. Formally, an aPSCF is a PCSF $F : \mathcal{B}(n, k) \to \mathcal{B}(k)$ given by $F(B) = (b_1, \ldots, b_k) = \vec{b}$ so that each $b_j = \frac{1}{n} \sum_{i \in N} b_{ij}$. The process of belief aggregation and collective decision making at some round t, can thus be given by the following algorithm.

Algorithm 2 Belief Aggregation at round t

Input: Belief matrix B^t
  1: for all $s(j) \in S$: $b_j^t := \frac{1}{n} \sum_{i \in N} b_{ij}^t$
  2: $\vec{b}^t := (b_1^t, \ldots, b_k^t)$
Output: Societal belief vector $\vec{b}^t$

Since an aPSCF does not rely on weights, it can be used as preference aggregation methods as long as an agreement on weights is not reached yet. Note that an aPSCF can actually be thought of as a special case of a wPSCF where the weight vector is given by $\vec{w} = (\frac{1}{n}, \ldots, \frac{1}{n})$. Therefore, an aPSCF satisfies all properties that the more general wPSCFs satisfy. In fact, the averaged method satisfies *non-dictatorship* (no single individual can always determine the social probabilities) and *consistency* (equal societal preferences of two disjoint groups are maintained when the two are merged).

Finally, let us consider *strategy-proofness* (SP), which indicates whether individuals can manipulate the outcome of the aggregation procedure when submitting an untruthful individual preference. It easy to verify that neither an aPSCF, nor a wPSCF satisfies SP. As we assume all players in the game to be group-rational, no player will manipulate the game with the purpose of increasing his private payoff only, so we do not worry about dishonest players trying to sabotage the co-operation. However, since the utility functions of the players are assumed to be unknown, they are not certain about the social optimum either. Hence, a manipulation by a very 'uninformed' player can be harmful for the entire coalition.[3]

4 Gameplay and Reinforcement Learning

Recall that the first belief update was performed after network communication, *before* the game starts. A second belief update is performed *after* playing the game. Whereas the first

---

[3]In Section 5, we will therefore introduce some conditions on the amount of influence of different individuals in the coalition, ensuring that such players will not have enough power for manipulation.

	IA	U	N	A	P	SR	ND	C	SP
wPSCF	✓	✓	✓	✓	✓	✓	-	-	-
aPSCF	✓	✓	✓	✓	✓	✓	✓	✓	-

Figure 1: Overview of axiomatic properties

update involves individual learning by way of communication, the second update involves collective learning by way of reinforcement. In a classical method for reinforcement learning [Bush and Mosteller, 1955] the probability for a certain strategy is updated with a weighted average of the previous probability and the maximum attainable probability 1. More specifically, suppose player i chooses the individual strategy s_i^t at round t and he receives a payoff of $u_i(s^t)$, where s^t denotes the joint strategy played at round t. Then the probability for playing s_i again, is increased by adding some *fraction* of the distance between the original probability for s_i and the maximum probability 1. This *fraction* is given by the product of the payoff and some learning parameter λ. The payoffs as well as the constant fraction λ are scaled to lie in the interval from 0 to 1. The size of λ correlates with the speed of learning [Skyrms, 2010]. The probabilities for all strategies that are not played in the previous rounds, are decreased proportionally.

Formally, let $m_i^t : S_i \to [0, 1]$ be the mixed strategy for player i at round t. Then, after playing s_i^t at round t, player i can revise his mixed strategy for the next round $t + 1$ as follows:

$$m_i^{t+1}(s_i) = \begin{cases} m_i^t(s_i) + \lambda \cdot u_i(s^t)(1 - m_i^t(s_i)) & \text{if } s_i = s_i^t \\ m_i^t(s_i) - \lambda \cdot u_i(s^t)(m_i^t(s_i)) & \text{if } s_i \neq s_i^t \end{cases}$$

Note that this reinforcement method is aimed to model how players can *individually* learn to improve their private strategy in repeated games. Let us extend this learning behavior to a group level, by allowing the coalition for reinforcing the joint strategies that yield a positive social welfare, i.e., sum of individual payoffs. In that way, players are *collectively* learning towards the social optimum.

More specifically, after belief aggregation by the aPSCF, the coalition holds a societal probability distribution over the set of joint strategies, which (for a round t) is presented by the stochastic vector $\vec{b}^t$. A joint strategy $s(j) \in S$ is now chosen to be played with probability b_j^t. Subsequently, players get to know the corresponding social welfare and calculate an average fraction $U(s^t) := \frac{1}{n} \sum_{i \in N} u_i(s^t)$. Instead of the individual payoffs, this fraction is used by each player for reinforcement, towards the joint strategy with a maximal social welfare. Although players perform this second belief update individually, as they are all reinforcing the same joint strategy with the same reinforcement factor, they are collectively learning towards the social optimum. Note that each player is actually learning which individual *role* to adopt in the team, i.e., which actions of his are consistent with the social optimum. This process of gameplay and reinforcement at a certain round t can be described by the following algorithm.

Algorithm 3 Gameplay and Reinforcement at round t

Input: Societal belief vector $\vec{b}^t$; Belief matrix B^t

1: $s^t := s(j)$, s.t. $s(j)$ is drawn from S with probability b_j^t
2: $U(s^t) := \frac{1}{n} \sum_{i \in N} u_i(s^t)$
3: for all $i \in N, s(j) \in S$:

$$b_{ij}^{t+1} := \begin{cases} b_{ij}^t + \lambda \cdot U(s^t)(1 - b_{ij}^t) & \text{if } s(j) = s^t \\ b_{ij}^t - \lambda \cdot U(s^t) b_{ij}^t & \text{if } s(j) \neq s^t \end{cases}$$

4: $B^{t+1} := (b_{ij}^{t+1})_{n \times k}$

Output: Probability matrix B^{t+1}

There are two main reasons for using the Bush-Mosteller reinforcement method rather than, for instance, the one based on *Polyá urns* [Erev and Roth, 1995]. Firstly, Bush-Mosteller reinforcement makes use of utility values that are scaled in the interval from 0 to 1. This guarantees that the utilities are in the same scale for all players, thus avoiding unequal influences of different players. Moreover, it ensures that the same unit is used for payoffs as for individual beliefs about the strategy profiles. Thus when reinforcing after gameplay, the utility values are appropriately used as some type of *weight* in order to update the beliefs.

Secondly, the Bush-Mosteller model does not take into account the accumulated rewards of earlier plays, so that the proportion of reinforcement does not get smaller over time. For modeling processes of collective learning rather than individual learning, it may make more sense to violate this principle (the so-called *Law of Practice*). Namely, the learning process depends on the different communication patterns in every round, so that it does not slow down.

5 The Game-Network Learning Model

In this section we combine the computational methods discussed in the previous three sections. The model, which we will call *Game-Network Learning Model*, describes an iterative process that follows the procedures of network communication, belief aggregation, and gameplay and reinforcement in each round (Figure 2).
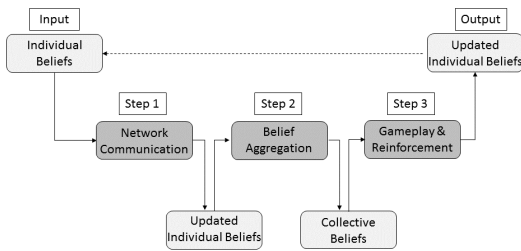

Figure 2: Main Loop of Game-Network Learning Model

Formally, the Game-Network Learning model thus consists of a run of Algorithms 1, 2, and 3. We will write B^{t^+} instead of B^{t+1} to denote the updated beliefs after network communication in round t. We emphasize that these are not yet the final updated beliefs at the end of round t, but merely the intermediate beliefs after communication. During each round t, the beliefs are thus updated twice, following the scheme:

$$B^t \quad \longrightarrow \quad B^{t^+} \quad \longrightarrow \quad B^{t+1}$$

Algorithm 1 thus outputs the matrix B^{t^+} after network communication in round t, which will subsequently be used as input for Algorithms 2 and 3.

5.1 Network Structure and Learning Outcome

Can adding network communication to a cooperative game be beneficial for the learning outcome? And if so, under what circumstances? Let us start with showing that network communication only influences the learning behavior of players as long as an *agreement* of beliefs is not reached yet. Formally, we say a group $N' \subseteq N$ is in *agreement at round t* if for all $i_1, i_2 \in N'$ it holds that $b_{i_1 \bullet}^t = b_{i_2 \bullet}^t$, i.e., if agents in N' hold the same probability distributions over S. Also, a group of nodes in a weighted directed graph is said to be *closed* if there are no outgoing edges from nodes inside the group to nodes outside the group. One could imagine that once everyone in a closed group in the social network has the same beliefs about the social optimum of the game, then agents of that group do no longer need to convince each other of their different opinions.

Proposition 1. *Let $N' \subseteq N$ be a closed group of agents in the network. Once N' is in agreement at the beginning of some round t in the Game-Network Learning Model, network communication in that round leaves the beliefs of all agents in N' unchanged, i.e., $b_{i \bullet}^{t^+} = b_{i \bullet}^t$ for all $i \in N'$.*

It follows immediately that for all agents $i_1, i_2 \in N'$ we find $b_{i_1 \bullet}^{t^+} = b_{i_1 \bullet}^t = b_{i_2 \bullet}^t = b_{i_2 \bullet}^{t^+}$ so that they will still be in agreement after network communication. Since the entire social network is always closed, once *all* agents agree, communication is no longer needed for an individual to learn towards the social optimum.

In order to show a *positive* influence of network communication on the learning process, we measure the quality of our Game-Network Learning Model by the probability for playing the social optimum at a given round (before agreement). More specifically, if $s(j^*) \in S$ is the social optimum, we say learning with network communication in round t is *better* than learning without network communication if $b_{j^*}^{t^+} > b_{j^*}^t$, where $b_{j^*}^{t^+} = \frac{1}{n} \sum_{i \in N} b_{ij^*}^{t^+}$ denotes the probability that the social optimum will be played in round t after network communication and $b_{j^*}^t = \frac{1}{n} \sum_{i \in N} b_{ij^*}^t$ denotes the probability that the social optimum will be played at round t without (or: before) network communication.

Intuitively, one could imagine that if there exist *experts* in the network, who are very close to knowing what the social optimum is, and these experts receive a sufficient amount of trust from all other players in the network, they can convince other players to increase the probability values for the social optimum. Formally, let $s(j^*)$ be the social optimum and let b_j^t be the probability that society assigns to some $s(j)$ at the beginning of round t. We say an agent $i_e \in N$ in the network is an *expert for round t* if $b_{i_e j^*}^t > b_{j^*}^t$. We write $\mathcal{E}^t$ for the

set of experts for round t. We call the agents $i \in N\backslash\mathcal{E}^t$ *non-experts*. Note that it follows directly that for all experts $i_e \in \mathcal{E}^t$ and all non-experts $i \in N\backslash\mathcal{E}^t$ it holds that $b^t_{i_e j^*} > b^t_{ij^*}$.

Intuitively, the experts for a certain round are the agents that have in the beginning of that round (and thus at the end of the previous round) a higher than average degree of belief for the social optimum. Note that experts can only exist as long as a total agreement is not reached. Namely, if an agreement of beliefs is reached between all agents in the network, every agent has the same degree of belief that is then trivially equal to the average degree of belief. The notion of an expert is therefore a *relative* rather than an *objective* one: an agent is only an expert when he has sufficient expertise relative to the expertise of others in the society.

Among the set of experts, there always exists a subset of experts who have the highest degrees of belief for the social optimum, compared to all other agents in society. These experts can be considered as *maximal* experts. Formally, if $s(j^*)$ is the social optimum, then we define the set of *maximal experts for round* t as $\mathcal{E}^t_{\max} = \{i_m \in \mathcal{E}^t \mid b^t_{i_m j^*} = \arg\max_{i \in N} b^t_{ij^*}\} \subseteq \mathcal{E}^t$. Note that it follows directly from this definition that for all maximal experts $i_m \in \mathcal{E}^t_{\max}$ and all other agents $i \in N\backslash\mathcal{E}^t_{\max}$ it holds that $b^t_{i_m j^*} > b^t_{ij^*}$.

Whether or not experts can exert more influence on the outcome of network communication than others, depends on their position in the network and the weight that other individuals assign to the opinions of the experts. To analyze this issue, we look at one's *centrality* [Jackson, 2008]. We introduce the notion of *weight centrality* to express the centrality of agents in weighted directed graphs. Formally, let $w_i = \sum_{m \in N} w_{mi}$ be the total weight that agent i receives from his neighbors. The *weight centrality* of some agent $i \in N$ is given by the fraction $C^w_i = w_i/n$. In words, weight centrality is a fraction of the highest possible weight that an agent i can receive, which is n (namely in case all agents would assign i a weight of 1, including agent i himself). The higher the weight centrality of experts, the more influence experts have on the outcome, and hence the higher the probability will be for playing the social optimum after network communication. The following theorem provides a sufficient condition for network communication to be *better* than no communication.

Theorem 1. *Let $s(j^*) \in S$ be the social optimum. If $C^w_{i_m} > \frac{1}{n} \geq C^w_i$ for all $i_m \in \mathcal{E}^t_{\max}$ and $i \in N\backslash\mathcal{E}^t_{\max}$, then the probability for playing the social optimum at round t after network communication is higher than before network communication, i.e., $b^{t^+}_{j^*} > b^t_{j^*}$.*

The theorem shows that if maximal experts for round t are trusted more than others, network communication can be beneficial for finding the social optimum. The proof relies on a decomposition of the individual weights into $w_{i_m} = 1 + \alpha$ for all maximal experts and into $w_i = 1 - \beta$ for all other agents (with α, β non-negative fractions). One can then show that the *increase* of the societal probability for the social optimum, due to communication with trusted experts, is greater than the respective *decrease* due to communication with less trusted non-experts. This guarantees that the probability for playing the social optimum at some round t after network communi-

cation is higher than before network communication.

If communication can be beneficial for a *single* round, we should ask if it can also be favorable in *every* round. We therefore introduce the notion of *stable experts*, which are the agents $i \in N$ for which it holds $i \in \mathcal{E}^t$ for every round $t \geq 1$. The following theorem states a sufficient condition for an initial expert (i.e., expert for round 1) to be a stable expert. In fact, the theorem states an even stronger result that ensures initial experts always to be in the set of *maximal* experts.

Theorem 2. *Let $s(j^*) \in S$ be the social optimum and let $\mathcal{E}^1$ be the set of initial experts. If $\mathcal{E}^1$ is closed and $\mathcal{E}^1$ is in agreement at round 1, then $\mathcal{E}^1 \subseteq \mathcal{E}^t_{\max}$ for all $t \geq 1$, as long as a total agreement in the network is not reached.*

Hence, if initial experts only assign positive weights to themselves or other initial experts with the same degrees of belief, then they will always be in the set of maximal experts for each round $t \geq 1$. The proof relies on Proposition 1 and the intuition that agents with maximal degrees of belief for the social optimum *after* network communication, are the agents with maximal degrees of belief for the social optimum after gameplay and reinforcement learning too.

In fact, the group of maximal experts might become bigger than the group of initial experts, but 'new' maximal experts can only have a degree of belief for the social optimum that is at most as high as the respective degree of belief of the initial experts. Under certain conditions of $\mathcal{E}^1$, it even holds that the group of initial experts is *exactly* the group of maximal experts for each round $t \geq 1$, i.e., $\mathcal{E}^1 = \mathcal{E}^t_{\max}$.

Proposition 2. *Let $s(j^*) \in S$ be the social optimum and let $\mathcal{E}^1$ be the set of initial experts for round $t = 1$. If $\mathcal{E}^1$ is maximally closed and $\mathcal{E}^1$ is in agreement at round $t = 1$, then for all $t \geq 1$ it holds that $\mathcal{E}^1 = \mathcal{E}^t_{\max}$, as long as a total agreement in the network is not reached.*

Here we call a group $M \subseteq N$ *maximally closed* if all agents outside of M are connected to at least one other agent outside of M. Intuitively, if the group of initial experts is maximally closed, then it means that for all agents $i \in N\backslash\mathcal{E}^1$ there must exist another agent $j \in N\backslash\mathcal{E}^1$ so that $w_{ij} > 0$. This guarantees that agents outside of $\mathcal{E}^1$ will always have a smaller degree of belief than agents inside $\mathcal{E}^1$ at every round $t \geq 1$, since the updated beliefs are weighted arithmetic means of beliefs of neighbors.

Corollary 1. *Let $s(j^*) \in S$ be the social optimum and let $\mathcal{E}^1$ be the set of initial experts. If (i) $\mathcal{E}^1$ is maximally closed; (ii) $\mathcal{E}^1$ is in agreement at round $t = 1$; and (iii) $C^w_{i_e} > \frac{1}{n} \geq C^w_i$ for each $i_e \in \mathcal{E}^1$ and $i \in N\backslash\mathcal{E}^1$, then $b^{t^+}_{j^*} > b^t_{j^*}$ at each round $t \geq 1$, as long as at total agreement in the network is not reached.*

In words, under the stated conditions for initial experts, the probability for playing the social optimum is in each round higher after network communication than before (or without) network communication. This corollary thus provides a sufficient condition for learning with network communication to be *better in the long run* than learning without network communication. The proof follows immediately from Theorem 1 and Proposition 2. Namely, from the assumptions (i) and (ii) it follows that the initial group of experts is always the group

of maximal experts, i.e., $\mathcal{E}^1 = \mathcal{E}^t_{\max} \subseteq \mathcal{E}^t$ for all rounds $t \geq 1$. Now if this group of maximal experts satisfies the stated condition for weight centrality (iii), it follows from Theorem 1 that the probability for playing the social optimum after network communication is higher than without communication in *every* round.

6 Conclusions and Outlook

In this paper we have studied the possibility of formal modeling the process of learning in game scenarios, among agents arranged in a social network. We proposed an iterative model of learning that follows the procedures of network communication, belief aggregation, and gameplay and reinforcement learning. We conclude that interaction in *specific* social networks can positively influence the learning behavior of players in a cooperative game. Learning with network communication can be better than learning without communication, when there exist players in the game who know better than average which joint strategy corresponds to the social optimum. If these players are sufficiently trusted by society, and players with little knowledge about the social optimum are considered less authorial, then the knowledgeable players can convince other players towards the social optimum.

We envision possible extensions of the model in the domain of competitive games. For example, one could make use of cooperative game theory to describe *transferable utility (TU) games*, in which several coalitions can play against each other. Our model could be extended to a setting in which different social networks, representing different coalitions, play a competitive game. Additionally, in our model we assume players not to know their payoff functions, and not to be aware of the entire network structure. By utilizing epistemic models, one could elaborate on these different notions of (restricted) knowledge by means of Dynamic Epistemic Logic. A reasonable amount of work on logics for social networks already has been carried out by (among others) [Christoff and Hansen, 2014; Liu *et al.*, 2014; Seligman *et al.*, 2013], and [Zollman, 2012], although not in the context of cooperative games.

Each component of our Game-Network Learning model is inspired by existing computational approaches. Varying the choices made for modeling each of them, would result in new learning paradigms. For example, instead of relying on Lehrer's and Wagner's model, one could use Bayesian learning, see [Bala and Goyal, 1998]. Moreover, instead of assuming static networks, one could explore a more dynamic setting in which agents arrive over time and are allowed to change their weights of trust. As for the preference aggregation procedure, one could adopt one of the strategy-proof probabilistic social choice functions proposed by [Barberà *et al.*, 1998]. For the procedure of gameplay and reinforcement learning, one could possibly allow for more than one social optimum, and rely on the *selfishness level* [Apt and Schäfer, 2014] as reinforcement fraction.

Finally, on the empirical side, we envision computer simulations in order to check the convergence properties of the proposed learning model. Also, testing our model by means of psychological experiments with human participants, would help shed some light on the use of the proposed learning paradigm in modeling real human learning in serious games.

Acknowledgments

References

[Apt and Schäfer, 2014] K.R. Apt and G. Schäfer. Selfishness Level of Strategic Games. *Journal of AI Research*, 49:207–240, 2014.

[Arrow, 1951] K. Arrow. *Social Choice and Individual Values*. John Wiley and Sons, New York, 1951.

[Bala and Goyal, 1998] V. Bala and S. Goyal. Learning from neighbors. *Review of Economic Studies*, 5:205–228, 1998.

[Barberà *et al.*, 1998] S. Barberà, A. Bogomolnaia, and H. van der Stel. Strategy-proof probabilistic rules for expected utility maximizers. *Mathematical Social Sciences*, 35:89–103, 1998.

[Bush and Mosteller, 1955] R. Bush and F. Mosteller. *Stochastic Models of Learning*. John Wiley and Sons, New York, 1955.

[Camerer and Ho, 1999] C. Camerer and T. Ho. Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4):827–874, 1999.

[Christoff and Hansen, 2014] Z. Christoff and J.U. Hansen. Dynamic Social Networks Logic. *Journal of Applied Logic (to appear), available as ILLC Prepublication Series Report PP-2014-09*, 2014.

[DeGroot, 1974] M.H. DeGroot. Reaching a Consensus. *American Statistical Association*, 69(345):118–121, 1974.

[Easly and Kleinberg, 2010] D. Easly and J. Kleinberg. *Networks, Crowds, and Markets*. Cambridge University Press, New York, 2010.

[Erev and Roth, 1995] I. Erev and A.E. Roth. Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term. *Games and Economic Behavior*, 212:164–212, 1995.

[Erev and Roth, 2014] I. Erev and A.E. Roth. Maximization, learning, and economic behavior. *Proceedings of the National Academy of Sciences*, 111, 2014.

[Fudenberg and Levine, 1998] D. Fudenberg and D.K. Levine. *The Theory of Learning in Games*. The MIT Press, Cambridge, 1998.

[Gibbard, 1977] A. Gibbard. Manipulation of schemes that mix voting with chance. *Econometrica*, 45(3):665–681, 1977.

[Intriligator, 1973] M.D. Intriligator. A Probabilistic Model of Social Choice. *The Review of Economic Studies*, 40(4):553–560, 1973.

[Jackson, 2008] M.O. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.

[Laslier *et al.*, 2001] J. Laslier, R. Topol, and B. Walliser. A Behavioral Learning Process in Games. *Games and Economic Behavior*, 37:340–366, 2001.

[Lehrer and Wagner, 1981] K. Lehrer and C. Wagner. *Rational Consensus in Science and Society*. D. Reidel Publishing Company, Dordrecht, 1981.

[Liu *et al.*, 2014] F. Liu, J. Seligman, and P. Girard. Logical dynamics of belief change in the community. *Synthese*, 191:2403–2431, 2014.

[Nitzan, 1975] S. Nitzan. Social Preference Ordering in a Probabilistic Voting Model. *Public Choice*, 24:93–100, 1975.

[Niv, 2009] Y. Niv. Reinforcement learning in the brain. *Journal of Mathematical Psychology*, 53:139–154, 2009.

[Seligman *et al.*, 2013] J. Seligman, F. Liu, and P. Girard. Facebook and the epistemic logic of friendship. In *Proceedings of Theoretical Aspects of Rationality and Knowledge, Chennai, India*, 2013.

[Skyrms, 2010] B. Skyrms. *Signals: Evolution, Learning & Information*. Oxford University Press, Oxford, 2010.

[Zollman, 2012] K.J.S. Zollman. Social network structure and the achievement of consensus. *Politics, Philosophy & Economics*, 11(1):26–44, 2012.