Emergence of Social Networks via Direct and Indirect Reciprocity

Washington Redskins

ebay®

	Fair Sale	Seller Scam
Fair Purchase	😃 😃	😃 😢
Buyer Scam	😢 😃	😢 😢

Prisoner’s Dilemma

Social Network

- Graph of agents
- Neighbors play a social dilemma game
  - One agent donates utility, bearing an initial cost
  - The recipient receives a multiplied amount of utility
- Agents learn which neighbors cooperate and which neighbors defect
- Alliances and coalitions emerge and disappear strategically

Agent Behavior

- Defecting is optimal in a single round
- Cooperation becomes most profitable in games played indefinitely
    - There is still incentive to defect strategically

- The agents need to learn which neighbors they can cooperate with

Direct Reciprocity

“agents condition their behaviour on personal experience of other agents in order to elicit cooperation”

Indirect Reciprocity

“being generous to strangers in order to gain a good reputation, thus allowing entry into profitable coalitions”

Previous Studies

- Created static networks (exogenous / top-down) and examined which parameter values led to collaboration among the agents
- Networks with small-world topologies, such as those created by preferential attachment, produced the most cooperation

Previous Studies (Direct Recip.)

- Some studies allow agents to connect to nearby agents and disconnect from others
- This allows for strategic manipulation of the network
- However, it does not support indirect reciprocation due to the localization of interactions

Previous Studies (Indirect Recip.)

- Studied networks are very large
  - More tractable to analytical techniques
  - Not typical in the real world
- The importance of the source of reputation information can be analyzed
  - Agents may trust their closer/stronger allies regarding the reputation of strangers, rather than trusting what strangers say about other strangers

Human Social Networks

- Highly dynamic at the individual level
  - Node degree
- Remain stable globally
  - Network diameter
  - Clustering coefficient

- Can’t fully be explained by direct reciprocity or indirect reciprocity alone

This study

- Agents are allowed to interact with all other agents
- The network emerges from individual interactions between agents (endogenous / bottom-up)
- Reputation information is conveyed through the resulting network

Model & Methodology

Katie Boylen

Portfolio

- Agents invest in partners
- Partners receive a multiple of the investment, m > 1
- Every agent has a portfolio of donations at each time step t

$$P^t_{i,*} = (w_1, w_2, \ldots w_n) \quad (1)$$

$$p^t_{i,j} \in [0, 1] \subset \mathbb{R} \ \forall_{i,j} \quad (2)$$

$$p_{i,i} = 0 \ \forall_i \quad (3)$$

$$\sum_{j=0}^{n} p^t_{i,j} \leq 1 \ \forall_i \quad (4)$$

- w1, w2 … wn are weights of the donation to agents a1, a2 … an
- The matrix of donations between agents at time t: $C^t = \gamma P^t$,
- The payoff to agent ai: $u^t_i = \sum_{j=1}^{n} m \cdot p^t_{j,i} - \sum_{k=1}^{n} p^t_{i,k}$.

Reputation

- Choosing not to invest or to only invest a little results in a bad reputation score $r_i^t \in [0, 1] \subset \mathbb{R}$ for an agent, represented by

$$r_i^t = \sum_{j=1}^{n} C_{i,j}^t$$

- And agent can donate based on other agent's reputations (indirect reciprocity) and the history of donations received from that agent (direct reciprocity)
- An exponential moving average is used to summarize the time series and weight more recent values more $\bar{c}_{i,j}^t = \max(\kappa,\ \alpha \cdot c_{i,j}^t + (1-\alpha) \cdot \bar{c}_{i,j}^{t-1})$ where $\kappa = \frac{\gamma \cdot \hat{m}}{4n}$

Reputation

- Visualize donation matrix as weighted directed graph
- Can be used to weight reputation of other agents based on their distance
- Factor in that information from direct sources may be more trustworthy
- $\bar{r}_i^t = \alpha \cdot r_i^t + (1-\alpha) \cdot \bar{r}_i^{t-1}$ does not factor network distance into the exponential moving average
- $\phi_{i,j}^t = \frac{\bar{r}_j^t}{d_{i,j}}$ does, it is the networked version of the reputation scores of the matrix Φ^t where di,j is the shortest path from i to j on the graph defined by C
- Agents can choose either form of measurement

Strategies

Four strategies

1. Cooperative strategy- agent donates the endowment equally among all agents

$$p_{i,j}^{t} = \frac{1}{n-1} \forall_{a_j \in A : j \neq i}$$

1. Defect strategy- agent accepts donations without any reciprocation

$$p_{i,j}^{t} = 0 \, \forall_{a_j \in A}$$

Strategies

3. Reputation-weighted strategy- agent distributes donations based on other agent's reputation

$$p_{i,j}^t = \frac{\bar{r}_{i,j}^{\,t-1}}{\sum \bar{\boldsymbol{R}}_{i,*}^{\,t-1}} \;\forall_{a_j \in A : j \neq i}$$

- Reputation-weighted networked strategy- agent distributes donations based on networked reputation scores

$$p_{i,j}^t = \frac{\phi_{i,j}^{\,t-1}}{\sum \boldsymbol{\Phi}_{i,*}^{\,t-1}} \;\forall_{a_j \in A : j \neq i}$$

4. Tit for Tat strategy- agent donates in proportion to the moving average of inward donations

$$p_{i,j}^t = \frac{\bar{c}_{j,i}^{\,t-1}}{\sum \bar{\boldsymbol{C}}_{*,i}^{\,t-1}}$$

Learning

- Agent uses a reinforcement learning algorithm that is based on Q-learning to select a strategy
- The agent tries out the different strategies and then uses the payoff values to estimate the expected payoff of each strategy
- Attempts to find greedy strategy- strategy with best long-term reward
- Payoff values depend on the state as well as the strategy chosen
- The state is the agent’s reputation
- Rounds reputation to one of five values: {0, 1/4 , 1/2 , 3/4 , 1}

Learning

- The estimated payoff values are held in a table of Q values
- Table updated based on the equation

$$\begin{aligned} Q_{i,t}(s_{i,t'}, \theta_{i,t'}) = \alpha \cdot [U_{i,t'} + \beta \cdot Q_{i,t}(s*_{i,t}, \theta_{i,t})] \\ + (1-\alpha) \cdot Q_{i,t'}(s_{i,t'}, \theta_{i,t'}) \end{aligned}$$

where si,t' is the strategy that agent ai played in period t −1, α is the learning-rate parameter, β is the discount parameter and s∗i,t is the greedy strategy of agent ai

- The equation is a discounted exponential moving average of historical payoff samples
- Recent payoffs are weighted more

Learning

- Trade-off between exploiting the greedy strategy and exploring to find a better one
- The exploration methods used are
  - Epsilon-greedy selection- chooses at random a strategy, if the strategy chosen is not the greedy strategy, it chooses at random again
  - Softmax- the probability of choosing strategy a at time t' is

$$P(s_{i,t'} = a) = \frac{\exp(Q_{i,t}(a, \theta_{i,t})/\tau)}{\sum_b e^{Q_{i,t}(b)/\tau}}$$

Learning

- Reinforcement learning models use theories of learning from cognitive psychology and explain the deviations from game theory seen with real subjects
- The learning-theoretic equilibria can be related to game-theoretic equilibria in certain cases

Methodology

- Strong reciprocators: agents initialized without learning, only use reputation-weighted strategy
- Minor fraction are strong reciprocators, rest use the learning algorithm

Methodology

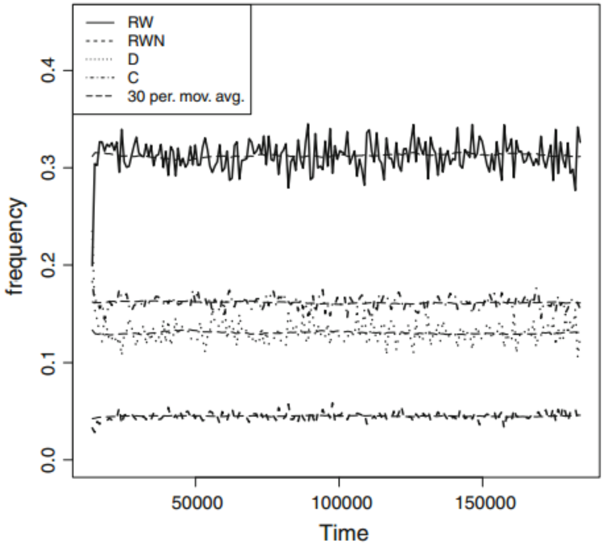
- 360,00 independent simulations were ran with these parameters

Table 1 Parameter settings

Parameter	Distribution	Description
ϵ	$\sim U(10^{-4}, 10^{-2})$	Experimentation
α	$\sim U(10^{-4}, 1-10^{4})$	Recency
β	$\sim U(0.9, 1-10^{4})$	Discount rate
Q_0	$\sim N(0, 100)$	Initial value estimate
n	$\in \{20, 60, 100\}$	Number of agents in the population
sr	$\in \{0,\ 0.05,\ \ldots,\ 0.4\}$	Proportion of strong reciprocators
m	$\in \{1.5,\ 2,\ 2.5,\ 3\}$	Multiplier

Methodology

The estimate of the level of cooperation in steady-state was taken to be the average reputation across the last 50,000 periods



Mean frequen a time series

Methodology

Study model when:

- learning is stateless and reputation does not factor into an agent’s choice of strategy
- learning is stateful and each agent’s reputation is used as a state value that factors into the agent’s strategy choice

Results

Trevor Poppen

Clarifications

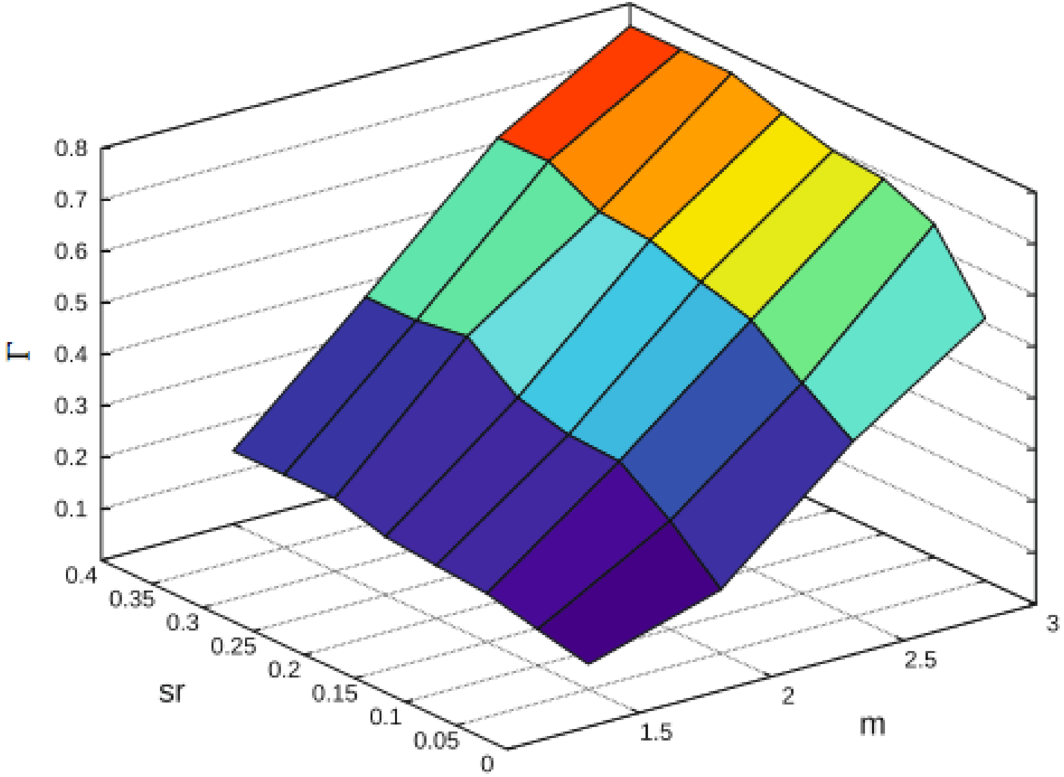
- Analysis is on steady-state simulations
- Time to equilibrium as not analyzed
- Solely conclusions and observations on equilibrium statistics

Stateless

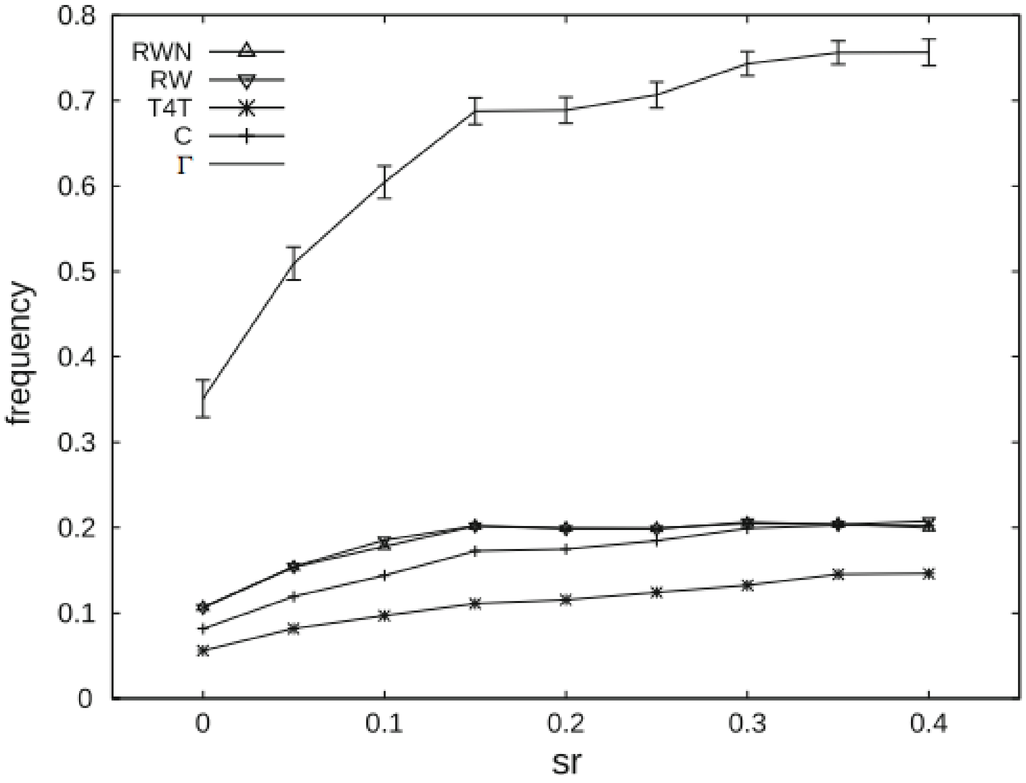
Regression fitting:

$$\Gamma = 0.29 \times m + 1.23 \times sr + 0.02 \times \beta - 0.44$$

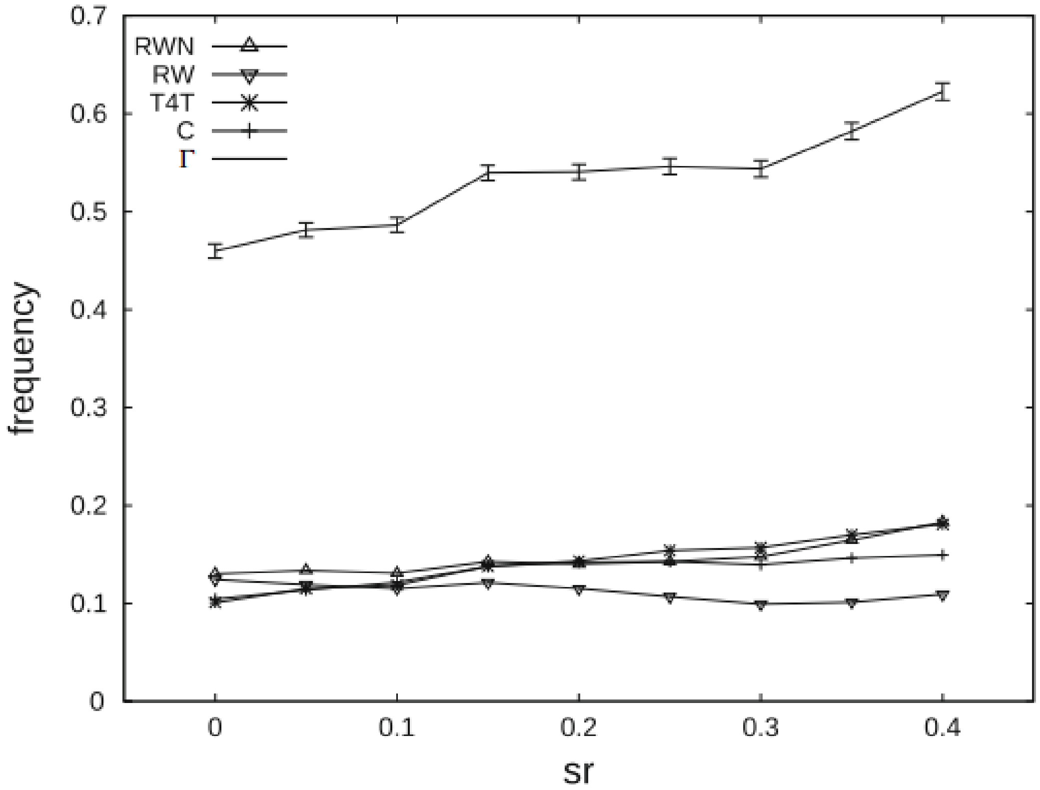
M,SR,Gamma



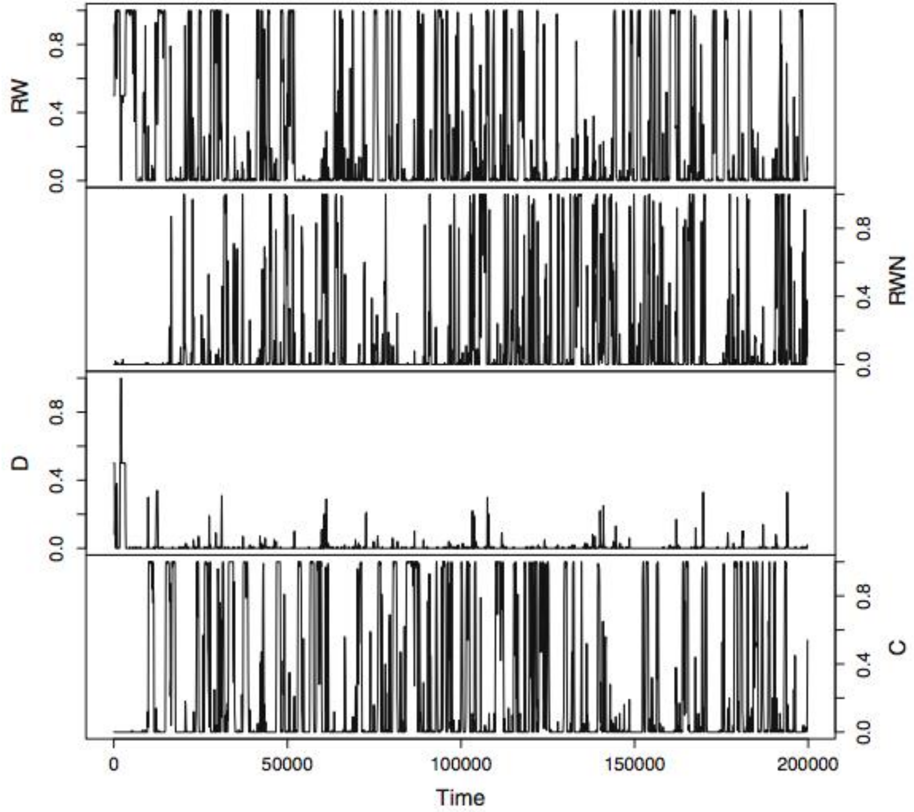
Stateless Strategy Contribution



Stateful Strategy Contribution



Individual Agent Behavior



Conclusion

Trevor Poppen

Key Contributions

- Both forms of reciprocity are important
- Interaction between both gives rise to networks which can reach equilibrium, but are still dynamic
- The differences of the two are direct results of the learning behavior

Outcome

- A network with a global equilibrium
- Agents with dynamic states
- Recency and Experimentation add dynamic behavior to environment
- Future work to be done with human subjects

Reference

Steve Phelps (2013). Emergence of Social Networks via Direct and Indirect Reciprocity, *Autonomous Agents and Multiagent Systems*, 27(3):355-374. (Phelps2013.pdf)

Questions?