

# Social Computing Article\*

## Abstract

Abstract

## Introduction

For years multi-agent systems have been used to research co-operation as a tool for problem-solving. Recently, however, there has been an increasing interest in the study of human beings as problem-solving agents. Several experiments have been conducted in which subjects are connected in a network with the goal of collectively solving a specific problem, and those have helped shed some light on the way humans interact to solve problems. However, although those studies can provide us with observations and hypotheses, there is still a difficulty in finding ways to explain the observed behaviour. In this paper, we use a multi-agent based simulation to complement the study of human computation, as a way of explaining the strategies used by humans and understanding their consequences in a cooperative environment. The control that simulations provide us over the behaviour of the agents allows us to better understand possible reasons for the ones displayed by humans.

Human beings are known to be able to easily perform tasks which are still generally difficult for computers, such as natural language processing and image recognition. However, it would be useful to be able to apply human social computation to more straightforward computer problems with precise definitions and algorithms, but which are still computationally intensive. The different sets of abilities between computers and humans suggest that the latter might provide a new approach to those problems that might be more effective than current techniques employed by computers. That notion has been successfully applied, for example, in one of the most significant problems in the field of bioinformatics: protein structure prediction (PSP). A software that presents the problem to humans as an online computer game, called *Foldit* (Cooper 2011), has produced significant results in the research of PSP, which is usually approached as an optimization problem requiring extensive computational power. Those results have been attributed to

human visual problem-solving and decision-making abilities, as well as social collaboration (Cooper 2011). However, we still don't know the limits of human abilities in problem-solving and how they compare to more traditional techniques. To take full advantage of human problem-solving strategies, we must learn their limitations. In order to obtain that knowledge, our system simulates human behaviour according to findings of social computation experiments.

## Background

Social computation is a relatively new area of research with a diversified, interdisciplinary root that mixes social sciences, artificial intelligence, game theory and network science, among others. Only recently studies on the potential of human social networks for solving problems are gaining some popularity. Those have provided insights into, among other things, the impact of network structure in the collaboration process and the factors that lead neighbours' proposed solutions to be copied by individuals.

The origin of human computation as we know it today can be traced back to the work of (Ahn and Dabbish 2008), which identified the possibility of using entertainment as an incentive to participation of human subjects, applying it in games in which the participants are actually performing a computation. That is an idea that also appears in the *Foldit* game.

The series of experiments summarized in (Kearns 2012) are among the first to try to take advantage of the collective's problem-solving abilities to solve classical computer problems. The experiments are mostly based on the concept of coordination: subjects have individual incentives that are expected to drive them to cooperate with one another and lead them toward the collective goal.

Other initiatives have appeared, such as the ones by (Farenzena *et al.* 2011) and (Mason and Watts 2011), which take a different approach by having subjects trying to solve the collective problem individually, with the possibility of exchanging solutions between neighbours. Those have resulted in interesting conclusions on human behaviour when the possibility of copying peers makes itself present. It's in those models of experiments that we will focus in this work.

The experiments conducted by (Farenzena *et al.* 2011) had human beings trying to solve constraint satisfaction problems, namely Boolean Satisfiability (SAT) and the pop-

---

\*This title is temporary

ular *Sudoku* game, with individuals connected through the network being able to exchange partial solutions of the problem in question. There were two main patterns of behaviour observed in this study in regard to copying neighbours' solutions:

- Humans might not evaluate the solutions proposed by their peers, instead choosing the most readily available one. On the other hand, (Mason and Watts 2011) suggest that their subjects did evaluate the available solutions, although an in-depth analysis of that wasn't provided.
- There is a higher probability of individuals copying peers' solutions when those are shared by several neighbours. That behaviour is referred to by the authors as *conformism*. A similar behaviour was observed by (Mason and Watts 2011), even though the experiments used a different problem.

(Mason and Watts 2011) also provided important insights to human cooperation. However, since our basis is the work of (Farenzena *et al.* 2011) we will use the former's results only as secondary support.

## Contribution

We have built a multi-agent system that simulates the experiments of (Farenzena *et al.* 2011), using the model of human behaviour proposed by them. We adapted that behaviour into a set of rules which we modeled in the *Memetic Network* model proposed by (Lamb and Araujo 2008). For simplicity, we limited our experiments to the *Sudoku* problem. Our goal is verifying whether the strategies employed by humans are competitive with other heuristics. After identifying the points in which those are lacking, we will then propose methods to increase the efficiency of human problem-solving networks.

## System Overview

We developed a system in which the environment of the *Sudoku* problem-solving social network presented in (Farenzena *et al.* 2011) could be modelled through autonomous agent networks and analysed through a series of experiments. The optimization technique of *Memetic Networks* (Lamb and Araujo 2008) was employed on the modelling, providing a basis for the dynamics of the agent network.

## Sudoku Solving Memetic Network

A *Memetic Network Algorithm* is composed of an ordered set of  $N$  agents, each encoding a complete solution to the optimization problem, and a binary  $N \times N$  matrix representing possible connections between agents. Additionally, a *Memetic Network Algorithm* is composed of a set of rules specifying how connections between agents are formed and erased and how interactions between these agents take place. These rules are grouped under the categories of *Connection Rules*, specifying how agents will connect and disconnect from each other; *Aggregation Rules*, specifying the dynamics of the information flow through connections; and *Appropriation Rules*, specifying how agents are supposed to add

local changes to the information received through their connections (Lamb and Araujo 2008).

Our solution adapts the *Memetic Network* technique - originally intended for use in optimization problems - to the context of constraint satisfaction problems. In this specific case (*Sudoku* solving), we treat problems with one unique solution. To deal with this scenario, we propose a *Memetic Network* variation in which each agent encodes not a complete, but a partial solution to the problem.

Our solution employs a fixed topology for the agent network and models cooperation through a set of *aggregation rules* and reasoning through a set of *appropriation rules*. Our aggregation rules specify how *Sudoku* partial solutions are copied from agent to agent, while our appropriation rules specify how agents increment these solutions with *Sudoku* solving techniques such as those studied in (Davis 2007).

## Aggregation Rules

The experiments conducted by (Farenzena *et al.* 2011) point out a series of observations about the dynamics of cooperation in problem solving with human beings. For instance, the authors analysis has shown that users tend to copy not the most complete solutions, but, instead, the first solutions on the graphic interface. Additionally, we know from this same paper that users rarely copy solutions from more than 6 neighbours. In order to analyse and compare this behaviours, we developed a collection of aggregation rules, each modelling a copying strategy.

**Pick Most Filled** This rule employs the intuitive strategy of copying from the neighbour with the most filled *Sudoku* board, namely the most complete solution.

**Pick Among First** We know that users tend to copy the first (from left to right) solutions on the graphic interface (Farenzena *et al.* 2011). The authors have provided us with a mathematical model of this behaviour, stated as  $\langle X(k) \rangle = (1 - p)^{k-1}p$ , where the parameter  $p$  is fixed as  $p = 0.5479$  and  $\langle X(k) \rangle$  denotes the probability of an agent copying the  $k_{th}$  neighbour solution.

We inserted this behaviour into our model by firstly generating a random ordering of neighbours for each agent in order to compose a simulated graphic interface. As a result, each agent visualizes some neighbours before or after others. Secondly, we translated the above mathematical model into an aggregation rule in which the solution copied by an agent is the  $k_{th}$  with probability  $\langle X(k) \rangle$ .

## Appropriation Rules

In *Sudoku* strategy literature, as in chess, we find multiple techniques with iconic names. Some popular examples are the *naked singles* rule, the *hidden singles* rule and the *naked twins* rule, studied in (Davis 2007). These strategies intend to, given a *Sudoku* puzzle in a partial state of completion, generate movements to mark blank cells of the puzzle, as "mark cell 1 of column 3 with value 5". We reproduced a variety of these strategies, modelling each one of them as a function that maps a *Sudoku* partial solution to a set of movements.

**Only Choice** This technique marks a *Sudoku* cell with a value only if this cell is the last blank cell on its row, column or  $3 \times 3$  block.

**Naked Singles** This rule marks a *Sudoku* cell with value  $v$  only if  $v$  is the single possible value to mark that cell. The other values must have been eliminated through a process of verifying that they are all present in the span encompassing the cell's row, column and  $3 \times 3$  block.

**Hidden Singles** This rule is similar to the *naked singles* rule. It is able to mark cells with more than one candidate value by looking for cells that are the only ones in their row, column or  $3 \times 3$  block that can hold a particular value.

**Two out of Three** This rule applies to groups of three contiguous  $3 \times 3$  blocks. It aims to find a value  $v$  such that  $v$  is present in two of the three  $3 \times 3$  blocks encompassed by the group, but missing on the third. It proceeds by enumerating all the candidate cells - namely the empty cells - on this block, and then, by eliminating from this set all the cells that are encompassed by the rows or columns in which  $v$  is placed on the other two blocks. If the resulting set has cardinality 1, the rule has successfully found a cell to mark.

### Naked Twins

### Copying Solutions

In a study conducted by (Farenzena *et al.* 2011), subjects were invited to solve *Sudoku* puzzles and share them with other subjects with whom they were connected through a network topology. At any time, a player could copy solutions from one of his or her neighbours. The authors' analysis has shown that there is a different copy frequency associated with each topology. For instance, the scale free topology with  $\gamma = 1.65$  resulted in a copy frequency of 87%, while the fully connected topology resulted in a copy frequency of 42%.

We incorporated this copying frequency into our modelling, recreating the scenario of the experiments conducted by (Farenzena *et al.* 2011) by setting this parameter accordingly to the topologies we used.

### Conformist Behaviour

We introduce in our modelling the *conformist* behaviour proposed by (Efferson *et al.* 2008) and further studied by (Farenzena *et al.* 2011), which states that agents tend to copy a particular solution with probability 1 when surrounded by neighbours that share this solution.

Our approach was to establish a *threshold* determining a minimum necessary quantity  $N$  of neighbours sharing a particular solution. In our modelling, all agents with  $N$  or more neighbours having a solution  $s$  in common are programmed to copy solution  $s$  from them.

### Guessing and Backtracking

We have consistent evidence that trial-and-error is a part of the *Sudoku* solving experience. The need for trial-and-error in *Sudoku* puzzles is not a falsifiable conclusion, but a mathematical fact (Davis 2007). Some puzzles are only solvable by the means of a backtracking procedure (see Figure 1).

|   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|
| 8 | 1 | 2 |   |   | 4 |   |   | 3 |
| 3 | 7 | 9 | 8 |   | 2 | 1 | 4 |   |
| 5 | 6 | 4 | 3 | 9 | 1 | 2 | 8 | 7 |
| 9 | 4 | 1 |   | 8 | 5 | 3 | 7 |   |
| 7 |   |   | 9 | 4 | 3 | 8 |   | 1 |
|   | 8 | 3 | 1 |   | 7 | 4 | 5 | 9 |
|   |   | 7 | 4 | 3 | 9 |   | 1 | 8 |
| 1 | 3 | 8 |   |   | 6 |   |   | 4 |
| 4 | 9 |   |   | 1 | 8 |   | 3 |   |

Figure 1: A *Sudoku* puzzle solvable only by trial-and-error

In our modelling, we associate each agent with a numerical parameter determining the probability of this agent to guess when incapable of applying a typical *Sudoku* solving strategy.

Automatic *Sudoku* solvers employing a backtracking algorithm are easily programmed and very time-efficient. On the other hand, the space complexity of these algorithms is a barrier to most human solvers, who need to write down tons of observations in order to employ a backtracking strategy. With this in mind, we propose in our modelling two different kinds of backtracking, intended to be more similar to the way human beings employ error correction in *Sudoku* solving in a collaborative environment such as the one analysed in (Farenzena *et al.* 2011).

- **Local Backtracking:** When faced with one or more conflicts in its own solution, an agent simply erases the conflicting cells.
- **Social Backtracking:** When faced with one or more conflicts in its own solution, an agent copies a solution from one of its neighbours. In our modelling, this is done by raising the copy rate of this particular agent to 1.

## Results

Results

## Conclusions

Conclusions

## Acknowledgements

Research partly supported by CPNPq and CAPES, Brazil.

## References

- Luis Von Ahn and Laura Dabbish. Designing games with a purpose. In *Communications of the ACM*, volume 51, pages 58–67, August 2008.
- Seth Cooper. Algorithm discovery by protein folding game players. June 2011.
- Tom Davis. The mathematics of sudoku. November 2007.

C. Efferson, R. Lalive, P.J. Richerson, R. McElreath, and M. Lubell. Conformists and mavericks: the empirics of frequency-dependent cultural transmission. In *Evolution and Human Behaviour*, volume 29, pages 56–64, 2008.

Daniel Farenzena, Ricardo Araujo, and Luis Lamb. Collaboration emergence in social networks with informational natural selection. In *2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing*, 2011.

Michael Kearns. Experiments in social computation. In *Communications of the ACM*, volume 55, pages 56–67, June 2012.

Luis Lamb and Ricardo Araujo. Memetic networks: analyzing the effects of network properties in multi-agent performance. In *AAAI'08*, pages 3–8, 2008.

Winter Mason and Duncan J. Watts. Collaborative learning in networks. November 2011.