Validating Agent-Based Models with Probabilistic Model Checkers

Alexia Balotescu
Technical Computer Science, University of Twente, Enschede, The Netherlands

ABSTRACT

This study explores the use of probabilistic model checking (PMC) to verify agent-based models (ABMs) from the social sciences. It focuses on a rumour-spreading model proposed by Mazzoli et al. (2018) [12]. The ABM has been encoded as a discrete-time Markov chain in the PRISM model checker. A set of 14 properties has been defined in order to formally verify the model. The results show consistency with the original paper's simulation results, confirming that formal verification can capture core dynamics, such as spontaneous activation of spreaders or debunking effects. However, challenges like state-space explosion or limited scalability demonstrate the trade-offs of this approach. This study shows that formal tools like PRISM are able to reproduce and even deepen the analysis of ABMs, offering formal guarantees and better insights into system behaviour. However, limitations were also observed, such as large memory usage and scalability issues. This positions PMC as a complementary method to simulation, valuable for medium-scale social networks with rule-driven or well-defined agent behaviours.

Keywords: Agent-Based Models, Rumour Spreading, Probabilistic Model Checking, PRISM model checker, Reproducibility Study, Social Simulation, Formal Verification

1 INTRODUCTION

1.1 Background and Context

Social sciences focus on studying human behaviour and general phenomena across domains such as politics, economics, and culture. Computer science is concerned with the design, analysis, and implementation of algorithms and systems that aid in facilitating life. These sciences are deeply intertwined as new technologies such as computational social sciences, are able to analyse social interactions [4]. One of the most recent and widely used approaches in linking these sciences is the Agent-Based Model (ABM). The agents represent entities which can make decisions autonomously [1]. Modelling such agents means observing their actions and decisions in a specific environment and drawing conclusions in order to simulate real-world situations and to assess the results. ABMs can bridge the gap between multiple disciplines, such as computer science and the social sciences, as they have the power to address problems from a large variety of sciences, as well as improve collaboration between different fields [2].

TScIT 43, July 4, 2025, Enschede, The Netherlands

When it comes to linking computer science to social sciences, although ABMs represent a powerful resource, they often lack validation of results and guarantees of the behaviours they simulate [7]. These challenges require methods capable of formally validating the correctness and reliability of model behaviours. Such ability can be achieved by using probabilistic model checking. Model checkers are tools able to validate the correctness of a given model. They are given two inputs: the system (represented by state transitions) and its formal property, and they output whether the model is indeed correct - and the property holds, or they get stuck due to complexity [9]. They perform this validation by analysing all possible scenarios of the model. Probabilistic model checkers (PMCs) extend this functionality, as they can return the probability of reaching a certain set of states (events).

ABMs heavily rely on empirical simulations to "understand" social phenomena. However, such understanding comes from simply observing the outcomes of many simulations, rather than having formal proofs [8]. Simulations alone are not able to capture all possible outcomes, especially considering rare or critical scenarios. PMCs can systematically explore all possible interactions and transitions between agents, therefore verifying whether the key properties are respected throughout the entire behavioural space.

This paper aims to conduct a reproducibility study on an agent-based model derived from social science literature. The scope is to formalize it for input into a probabilistic model checker (PRISM) [11], assess how accurately the model can be reproduced, analyse the resulting formal verification outcomes, and reflect on the findings and challenges that were encountered during the process. Overall, the purpose is to formally analyse the diffusion dynamics of the social science ABM using probabilistic model checking, in order to assess how reliably key outcomes can be captured beyond stochastic simulation.

1.2 Problem Statement

One of today's most pressing issues is the spread of fake news and misinformation, which has intensified with the emergence of social media and communication platforms. Different rumour-based models have been developed for decades, in order to assess the way in which rumours propagate in social networks [14]. Rumour spreading has been studied through two main approaches in modelling: macroscopic (based on mean-field equations) and microscopic (agent-based, where individual nodes interact locally over a network structure) [13].

In this research project, the agent-based model of rumour spreading proposed by Mazzoli et al. (2018) [12], is selected here for reproduction and formal verification. The paper presents a structured and extensively analysed rumour diffusion ABM, making it suitable for formal replication. Their model simulates the spread of information over a scale-free network in which agents are assigned individual scepticism thresholds (their likelihood of spreading rumours) [12].

They can become spreaders using three mechanisms: spontaneously sharing the news after being exposed to it, being socially influenced by neighbours, or being persuaded into sharing through direct communication. Their model captures the dynamic evolution of rumour propagation at the individual level and incorporates realistic social network structures. Nevertheless, their analysis is based on stochastic simulations rather than formally validating possible behavioural trajectories of agents in the system, which this paper addresses through formal analysis.

1.3 Research Questions

The above problem statement can be expressed through the following proposed research question:

To what extent can probabilistic model checking be used to verify key behavioural outcomes, such as complete rumour spread, early extinction, and correction effects, in the agent-based model derived from Mazzoli et al. (2018) [12] and implemented in PRISM?

This main research question is formulated for the purpose of this study and can be decomposed into the following sub-questions:

1. What is the probability that the rumour reaches all agents before it dies out?
2. How does the initial number of spreaders affect the probability of full diffusion?
3. How do the spreading mechanisms (spontaneous, persuasion, debunking) influence the system outcomes?
4. To what extent is PRISM a suitable tool for the formal analysis of such ABMs?

To answer these questions, the model described by Mazzoli et al. (2018) [12] was manually implemented as a Discrete-Time Markov Chain (DTMC) in the PRISM model checker. The implementation contains the original mechanisms described in the paper: spontaneous spreading, social influence, persuasion and debunking. Furthermore, several logical properties were defined to observe the different diffusion outcomes. These were verified using model checking and statistical simulation across multiple model instances with different agent counts, as PRISM supports both model checking and stochastic simulation. Each sub-question is mapped to one or more of the formal properties defined.

1.4 Paper Structure

The paper is structured into six main sections, each addressing a distinct part of the research. The first one introduced the background on agent-based modelling, probabilistic verification and the research motivations and questions. Section 2 provides a literature review, highlighting the work related to formalizing ABMs, rumour dynamics and the capabilities model checking tools such as PRISM. Section 3 explores the methodology of the research, including the formalization of the original model into a DTMC, the encoding of agents' behaviours and the properties written for formal verification. Section 4 showcases the results that are grouped by outcome type and theme, and analysed across the different agent configurations. Section 5 presents the discussion, elaborating on the implications of the findings, comparison to existing literature, evaluation of PRISM as a tool, and reflection on the study's limitations and potential extensions. Lastly, section 6 concludes the paper by summarizing the key contributions, answers the research questions and proposes directions for further work in integrating formal verification in computational social science.

2 RELATED WORK

Research into ABMs has been well established. Bravo and Farjam [4] provide an overview of the prospects and challenges for computational social sciences. Similarly, Abar et al. [1] provide a survey on the current ABM tools and environments, while Axelrod [2] discusses how such models can bridge the gap between theoretical modelling and empirical social research by simulating the behaviour of individual agents in complex environments. However, ABMs also face limitations, which arise from conceptual modelling challenges as well as technical constraints, such as the computational complexity caused by simulating autonomous decision-making or verifying emergent behaviours. Conte and Paolucci [6] address the limitations present in the realism of agents, particularly the difficulty of differentiating between agent- and environment-driven behaviour. Chopra et al. [5] explore the limits of agency in traditional ABMs and propose frameworks that could allow agents' autonomy to proportionally increase with system complexity while also maintaining realistic behavioural constraints.

Meanwhile, formal verification methods have gained popularity: tools like STORM (Hensel et al. [9]) and PRISM (Kwiatkowska et al. [10]) allow for an exhaustive analysis of probabilistic systems. PRISM and similar tools express the dynamic of a system as probabilistic state-transitions, most commonly Discrete-Time Markov Chains (DTMCs) [10]. They model systems that progress in discrete time steps, with the current state influencing the probability of transitioning to a new state. Banisch et al. (2011) [3] show how ABMs can be transformed formally into DTMCs and analysed for transient behaviour and absorbing states.

In the domain of rumour spreading, Nekovee et al. [13] model the diffusion of information on complex networks using mathematical approximations. Mazzoli et al. (2018) [12] extend this model by proposing an ABM where agents have individual scepticism thresholds influencing their behaviour. In this project, the focus will be on reproducing the model of Mazzoli et al. (2018) [12] and formally verifying its diffusion properties using PMC techniques.

3 METHODOLOGIES

This section presents the formal modelling steps taken in order to replicate and verify the rumour-spreading ABM.

3.1 Model Implementation

The first step in this research was to reproduce the rumour spreading model described above. In their model, Mazzoli et al. (2018) [12] place agents on a scale-free network, with each having individual scepticism thresholds that determine their chances of spreading information. There are three ways in which agents can become spreaders: spontaneous exposure to the news (such as encountering it on a timeline), being influenced by neighbours who have already shared the rumour, and persuasion through communication with another agent. Moreover, the paper includes a debunking mechanism which reduces the chances of further spreading, when misinformation is identified.

As model checking large-scale networks is limited, the original network structure was not directly preserved. Instead, a simplified system was adopted, in which agents do not interact over a network graph. Rather than modelling connections through a graph topology (such as nodes representing agents and edges representing connections), the interaction is abstracted by encoding influence as global conditions. All agents are modelled as isolated modules, with their interactions being encoded through global conditions representing influence. This abstraction preserves the essential behaviour dynamics from the original paper, while also allowing formal verification in prism. Experiments were conducted on models of 1,3,6,8 and 10 agents. In each configuration (model), agents were assigned individual thresholds and behaviour types, to simulate the heterogeneous scepticism from the original model. The probabilistic transition rules from the original work were respected; modifications occurred only when accommodating PRISM's syntax and execution model.

Throughout the process of implementing the model, emphasis was placed on maintaining fidelity to the original dynamics. The assumptions made on the core behaviour of agents were kept intact. Simplifications were made only in structural aspects which would otherwise lead to intractable state explosion.

The PRISM code for the different numbered agents and the property specification are accessible via this link: https://github.com/alexiamaria3/Probabilistic-Model-Checking-ABM

3.2 Model Formalization

The original rumour spreading model was formalized as a DTMC in the PRISM model checker. A DTMC, being a probabilistic system, allows for representing the stochastic interactions and state changes of agents, as described in the original ABM.

3.2.1 States. Following the principles described in the original paper, in PRISM each agent was represented as an individual module which contains a state variable, namely sX, with X being the index of the agent. The state space is encoded as an integer variable with the following mapping:

- 0: Ignorant (initial state, unaware of the rumour)
- 1: Spontaneous spreader (spreads without social exposure)
- 2: Influenced spreader (spreads due to neighbours' influence)
- 3: Persuaded spreader (spread after direct peer communication)
- 4: Stifler (no longer spreading)
- 5: Debunker (actively stopping the spread)

This categorization respects the roles and transitions of the original ABM, where the transitions of agents depend on both personal thresholds and peer behaviour.

3.2.2 Formulas. Constants and Boolean formulas were defined within the PRISM model to reflect the paper's scepticism thresholds and network-based influence. Each agent has a static threshold value specific to the agent's index (th0, th1 etc.) which is compared to the global reliability variable for spontaneous spreading. The influence of neighbouring agents is captured in formulas such as enough_inflX, X representing the respective agent, which activates when a minimum number of neighbours are in the spreading state.

3.2.3 Probabilistic Transitions. The transitions described by Mazzoli et al. were represented using probabilistic commands. For instance, spontaneous spreading occurs when an agent with a low scepticism threshold is exposed to news that carries a certain reliability. In the formal model, such behaviour is captured by guards – an example can be seen in the figure below. 'r' represents the global reliability score, and the probabilities reflect the chances of adopting the rumour.

```
// Spontaneous spreading
[] s0=0 & r >= th0 -> 0.3 : (s0'=1) + 0.7 : (s0'=0);
```

Influence spreading was encoded as a probabilistic transition which is activated when an agent has at least one neighbour actively spreading. The transition assigns an 80% chance of entering the influenced state (state 2), thus maintaining consistency with the stochastic assumptions in the original ABM. Persuasion is implemented as a separate mechanism: a non-spreading agent directly interacts with a neighbour who is already spreading. If the agents have similar scepticism thresholds, persuasion has a 70% success rate, pushing the agent into the persuaded state (state 3). These separate transitions capture the conceptual distinction between passive social influence and active peer communication.

```
// Influence spreading
[] s0=0 & enough_infl0 -> 0.8 : (s0'=2) + 0.2 : (s0'=0);

// Persuasion
[] s0=0 & is_spreading1 -> 0.7 : (s0'=3) + 0.3 : (s0'=0);
```

The debunking mechanism, where agents reject or revert spread, is implemented by assigned a special state, namely 5, which halts further spreading and can influence others. This behaviour was added in an extended version of the original model, to support verification of the correction mechanisms.

```
// Debunking (if another agent is a debunker)
[] s1=2 & s2=5 -> 0.6 : (s1'=0) + 0.4 : (s1'=2);
[] s1=2 & s4=5 -> 0.6 : (s1'=0) + 0.4 : (s1'=2);
```

3.2.4 Time and Rumour Reliability. Mazzoli et al. include time dynamics; the rumour credibility decreases after a fixed number of timesteps. This behaviour was modelled by creating a separate time module, with variables t (time) and r (rumour reliability). Once the tick count reaches a set threshold, r was lowered to simulate how scepticism grows over time.

```
// Time counter (t) to change reliability (r) later
module Time
    t : [0..100] init 0;
    r : [0..100] init 99;

    // tick every step, increase time by 1
    // if time reaches 5, update reliability (news becomes false)
    [tick] t < 100 -> 1.0 : (t'=t+1) & (r'=(t=5 ? 48 : r));
endmodule
```

3.2.5 Conditions and Rewards. In order to track the number of steps until termination, reward structures were used, which help assess properties such as "expected time until spread ends (all agents reach terminal states)". PRISM labels were defined to monitor the key behaviours for temporal formal verification:

- "spontaneous": any agent in state 1
- "persuaded": any agent in state 3
- "no_ignorants": all agents are non-zero
- "all_done": all agents either stiflers or debunkers
- "all_influenced": all agents are in state 2

Overall, this PRISM formalization captures the core behavioural transitions and agents interactions of the original paper, as well as abstracting the network structure to allow proper verification. The result is a reproductible model which allows for probabilistic model checking of key rumour diffusion behaviours.

3.3 Property Specification

After formalizing the model, the next step was to define a set of logical properties which can capture the essential behaviours of the system, considering the research questions. These properties are written in Probabilistic Computation Tree Logic (PCTL), a temporal logic which is supported by PRISM and can assess reasoning about stochastic processes such as DTMCs. Each property was designed to correspond to a key outcome of the rumour spreading process, allowing the probabilistic dynamics to be formally verified and compared across multiple scenarios. Each of the 5 models (1, 3, 6, 8, 10 agents) had a list of properties to be verified, which are categorized and explained below (all written examples are taken from the properties file of the model of 6 agents).

3.3.1 Final Spread Completion. This property calculates the probability of the rumour eventually stopping spreading, meaning that all agents have transitioned to a terminal state such as stifler or debunker (states 4 or 5). It checks all the possible agents (s0 to sn, n being the number of agents) having a final state greater or equal to 4. This answers the first research question by quantifying the likelihood of full diffusion. Formally, it tests whether the system always reaches a terminal absorbing configuration, essential in checking system convergence.

```
P=? [ F (s0>=4 & s1>=4 & s2>=4 & s4>=4 & s5>=4)]
```

3.3.2 Full Influenced Coverage. The property verifies if all agents enter the influenced spreader state by peer exposure only, capturing the outcome where everyone reaches the 'influenced' state (2) due to peer exposure only. This helps to check the dominance of the influence mechanism, showing how strong peer influence is across runs.

```
P=? [ F "all_influenced"]
```

3.3.3 Complete Awareness. With this property, it is determined if all agents leave the ignorant state, regardless of whether they spread the rumour. This verifies if the system covers the full agent space, thus validating the questions of effectiveness.

```
P=? [ F "no_ignorants"]
```

3.3.4 Spontaneous Spread Occurrence. This label tracks whether at least one agent spontaneously encountered the rumour based on its high reliability. It captures the influence of the initial reliability and how rumours are adopted autonomously, thus linking to the realism of spontaneous spreading and its impact on the system dynamics.

```
P=? [ F "spontaneous"]
```

3.3.5 Persuasion Effectiveness. This property observes if any agent became a spreader by being persuaded. It isolates the effect of one specific mechanism in the model (persuasion in this case), helping to understand the internal dynamics and verify assumptions made in the original ABM.

```
P=? [ F "persuaded"]
```

3.3.6 Debunking Effectiveness. With this label it is checked whether at least one agent was reverted by a debunker. This makes it possible to test the impact of the debunking mechanism, especially in the presence of scepticism. It adds realism to the social element of the model and also links to the fourth research question.

```
P=? [ F "debunked"]
```

3.3.7 Combined Trigger Conditions. This property tracks if the rumour was triggered by at least one of the spreading mechanisms. This is important for establishing how complete the model is, meaning that the transitions collectively lead to the rumour being activated. If this fails, it can indicate modelling or logic flaws.

```
P=? [ F ("spontaneous" | "persuaded" | "all_influenced")]
```

3.3.8 Time-Bounded Diffusion. This property measures the probability that all agents hear about the rumour within a timeframe of 10 steps. It provides insight into the short-term effectiveness of the rumour dynamics. Moreover, it highlights how fast the diffusion can occur, showcasing temporal patterns that could be missed in standard simulation.

```
P=? [ F<=10 "no_ignorants"]
```

3.3.9 Time-Bounded Termination. This tests whether all spreading stops within 15 steps, number chosen for limiting the expected number of steps while also allowing enough time for termination. It is useful for performance analysis, by checking if the model's termination can be guaranteed within a specific timeframe. High values could signal abrupt decay or high correction rates.

```
P=? [ F<=15 (s0>=4 & s1>=4 & s2>=4 & s4>=4 & s5>=4)]
```

3.3.10 Early Debunking. A time-bounded property which checks if any debunking happens within the first 5 steps. It is important in modelling how real-world misinformation corrects, especially in fast moving social networks. The number of 5 steps was chosen to strike a balance between responsiveness and realism. It is early enough to indicate responsiveness while also allowing time for interactions between a debunker and an agent.

```
P=? [ F<=5 "debunked"]
```

3.3.11 Deadlock Detection. This is a sanity check in formal verification which validates if the transition structure allows for total system exploration, ensuring that the system never reaches a state from which no transitions are possible.

```
P<1 [ F true]
```

3.3.12 State Space Size. Counter of the total number of reachable states in the model, used to quantify the complexity for the different agent counts.

```
filter(count,true)
```

3.4 Verification and Analysis

Once the model was formalized and the properties were specified, verification was conducted using the model checker in PRISM. For the smaller models of 1, 3 and 6 agents, formal model checking was used, where the full state space could be generated and explored.

For each of the three models, the PCTL properties listed in the above section were added to the PRISM properties tab for verification and tracking. Reward structures were used to compute the expected values, and state labels allowed for high-level behavioural queries. Each of the 14 properties was verified, and the Log tab within PRISM allowed for every model to have a separate file containing the outcomes of the verification. As the number of agents increased, the state space grew exponentially, reaching over 38 million states for the 8-agent model, and nearly one billion for the 10-agent version. Thus, due to the limitations of properly checking the model at this scale, as well as the exhaustive amount of time necessary to compute such formal verification, PRISM's statistical simulation was used for the 2 larger models. The simulations were run with 10000 samples, a confidence level of 99% and a maximum path length of 10000. These parameters allowed for very accurate simulation results alongside runtime feasibility.

Each property was evaluated on all model sizes. For example, the probability of full spread was constantly above 99% in all configurations. The time bounded properties allowed for direct comparisons with the simulation results in the original paper. Moreover, in the cases where simulation produced timelines spread probabilistically, the model checker showcased exact convergence probabilities and expected step counts. In contrast, scenarios such as "all_influenced" or "persuaded" gave mechanism-specific insights which can be difficult to isolate using stochastic simulation alone. During verification, deadlocks were also detected and resolved through model inspection. The model showed consistency in converging to terminal states, and rare paths were still accounted for due to the full-state analysis.

The analysis confirmed that, to some extent, PRISM can be capable of verifying high-level outcomes, as well as fine-grained agent behaviours. Properties were made time-bounded, mechanism-specific and reward-based, thus allowing a better and more complete interpretation of the diffusion dynamics compared to using simulation alone.

4 RESULTS

4.1 Overview

This section presents the outcomes of the formal verification process, conducted on the rumour-spreading model. Using PRISM, a series of properties were evaluated across multiple agent counts. These results are grouped thematically considering the nature of the property being tested.

4.2 Property-Based Analysis

Each subsection describes the results of running the property verification or simulation, grouped by the general information they represent. Each table will contain the appropriate values, with each property being referred to as the title from Section 3.3.

4.2.1 Full Diffusion Outcomes. These are the first three properties (3.3.1, 3.3.2, 3.3.3), which measure whether the rumour fully spreads or reaches all agents. These align with the first research sub-question, as they verify how often the system leads to full awareness or absorption.

Agent count	Final Spread	Full Influence	Complete Awareness
1	0.792	0	0.793
3	0.983	0	0.983
6	0.996	0	0.996
8	0.999	0	0.998
10	0.999	0	0.999

The results show that the diffusion coverage increases as the number of agents does, which suggests the model's mechanisms are very effective in pushing the system toward terminal states. The complete awareness property (checking if all agents leave the ignorant state) behaves similarly, since it reaches almost certainty from 6 agents onwards. In contrast, the full influence property is 0 for all model sizes, suggesting that agents rarely pass through a distinct "influenced" phase before becoming spreaders or stiflers. However, they do eventually engage with the rumour. In this formal model, agents seem to transition directly from ignorance to spreading without pausing in a separate, influenced state. This shows a modelling simplification not explicitly mentioned in the original paper, where influenced was implicitly considered via thresholds or neighbour pressure, rather than separately encoded. Thus, this model reveals how simulation captures widespread interaction but can also overlook intermediary states.

These results are a strong formal support for the original model's conclusions regarding the effectiveness of social exposure mechanisms in scale-free networks. The paper explored how the rumour penetration under certain conditions/thresholds is almost total; the formal model confirms this with probabilistic certainty. Nevertheless, the fact that "influence" alone is never reached without spreading indicates the fact that the model does not clearly distinguish between belief acquisition and information propagation; instead, it combines them into a unified process during formalization.

4.2.2 Mechanism-Specific Activation. The next category are the properties which track the role of different spreading parts (3.3.4, 3.3.5, 3.3.6, 3.3.7). They assess how each individual mechanism behaves, supporting the third research sub-question and validating internal logic.

Agent Count	Spontaneous Spread	Persuasion	Debunking	Combined Trigger
1	0.792	0	0	0.792
3	0.998	0.28	1	0.998
6	0.998	0.911	1	0.998
8	0.999	0.966	1	0.999
10	0.999	0.987	1	0.999

The values in the table show that spreading spontaneously is consistently active, with its probability reaching almost 1 for 3 or more agents. This confirms that the model's mechanism of threshold visualization (where an agent shares the rumour upon seeing it) is reliable and has consistent behaviour. It also emphasizes the importance of spontaneous activation being the initial trigger in spreading the rumour. However, the persuasion mechanism is different. It starts with a probability of 0 (since no peers are available to persuade) and rises from 0.28 (3 agents) to 0.987 (10 agents). This reveals how persuasion needs network interaction: a spreader must have neighbours who are active *and* meet the scepticism threshold. As the number of agents increases, the probability of such interactions naturally rises as well. These findings validate the existence of persuasion operation in the formal model, but its effectiveness becomes significant only in networks of moderate sizes. This aligns with the interpretation by Mazzoli et al. (2018) [12], where it was shown how peer communication is important in sustaining the rumour diffusion after the initial spread. The debunking mechanism is also confirmed to be active in every instance beyond 1 agent, with a perfect success rate (1). This perfect score confirms the conclusion that, once a debunker is present, its influence of correcting is certain to appear. It is interesting to note that this is independent of the agent count, so the debunking transition is dominant: whenever the conditions of reversion are met, it consistently succeeds. Thus, the authors' claim that sceptical agents are essential for tempering spread is confirmed. Lastly, the combined trigger property, acting like a logical consistency check (true is the rumour is activated by at least one mechanism) is consistent with the values of the spontaneous spread. This implies that spontaneous visualization is the initiating condition in nearly all cases, especially in smaller networks.

These findings strengthen the argument that each mechanism contributes in a complementary manner: spontaneity initiates the spread, persuasion sustains it and debunking tempers its growth. The probabilistic results clarify the specific thresholds at which the persuasion starts being impactful.

4.2.3 Time-Bounded Dynamics. These are the properties that check how fast certain actions happen (3.3.8, 3.3.9, 3.3.10). The purpose is to visualize the speed of diffusion and correction, important for practical effectiveness.

Agent Count	Diffusion	Termination	Early Debunking
1	0.766	0.524	0
3	0.981	0.684	1
6	0.088	<0.001	1
8	0.005	0	1
10	<0.001	0	1

The first column represents the probability of all agents reaching a state of awareness (no ignorants) within 10 steps. There is an inverted trend, as the first two instances present a fast and likely diffusion, while for larger agents the probability collapses, reaching virtually 0 for 10 agents. Such sharp decline reflects how growing complexity and branching affects diffusion, as more agents are added. Larger networks need more steps to achieve complete reachability, so time bounds like 10 steps are insufficient. This results helps to quantify how larger networks require more steps to reach full propagation, thus reinforcing the limitation of stochastic ABMs, namely slowing down under growing topologies. In contrast, the termination property (rumour dying within 15 steps) has a slightly different pattern. In small networks termination is likely but becomes nearly impossible for larger models, since from 6 agents onward the probability is virtually 0. As such, it is suggested that higher agent counts means the system remains "alive" longer and becomes more stable (or self-sustaining) through the reinforcement of peer interactions. This aligns with the observation in the original paper stating that diffusion tends to become persistent in scale-free networks, where hub notes reinject momentum into the system. The early debunking property remains stable (1) for all networks apart from the smallest one. This means that, as soon as the debunker is present, their intervention is quick and reliable, within 5 steps and regardless of network size. This validates the model's internal design in which scepticism acts as an immediate override; moreover, it supports the principle that even in large, dynamic environments, a single critical entity (debunker) can rapidly trigger correction.

4.2.4 Model Complexity. The last category represents the sanity checks of the integrity and state space of the model (3.3.11, 3.3.12), thus validating its correctness and scalability.

Agent Count	Deadlock Detection	State Space Size
1	false	303
3	false	1616
6	false	1504597
8	false	38790565
10	false	980364277

The results of the deadlock detection property, checking if there are reachable states in which no further transitions are possible (deadlocks) is constantly false in all instances, which is the expected outcome. This means that the system is always guaranteed to have a valid transition from every reachable state, thus confirming there are no premature halts or modules constructed improperly. It also validates the construction of the transition guards, so there is completeness in all behavioural paths under the PRISM syntax. The state space counter shows a predictable exponential growth as the number of agents increases. The first models' values are relatively small; however, there is a dramatic jump from the 6-agent model, which reaches 1.5 million states, and by 10 agents there are 980 million. This growth represents a well-known challenge in PMC: the model, although logically correct, presents a state-space explosion. As such, the model's scalability is limited, at least for exhaustive verification.

Interestingly, these findings are in accord to the limitations in the original paper, where it is stated that simulating large-scale networks with agent heterogeneity is computationally infeasible, beyond a certain threshold. In formal verification it appears to be around 6-8 agents; however, unlike stochastic simulations, model checking gives explicit confirmations of the structural validity or deadlock-free status.

4.3 Overall Observations

The findings showcased in the previous sections provide some important insights into the behaviour and formal structure of the replicated model, as well as how suitable it is for verification using probabilistic model checking.

Firstly, the model's outcomes become more stable as network size increases. In almost all categories, the probability of desirable outcomes, such as full diffusion, activating mechanisms or early debunking, tends to approach 1 as the number of agents increases. This trend is especially visible in properties like spontaneous spreading or persuasion, where values stabilize at almost 100% for the networks of more than 6 agents. Therefore, the system's internal logic becomes more stable when populations are larger, likely due to the likelihood of peer reinforcement increasing, as well as the diffusion redundancy present in a scale-free topology. At the same time, the results also show that long-term convergence becomes increasingly unlikely, as the model often fails to terminate or reach absorbing states within an explicit number of steps. This is visible especially in the low (or zero) values in the time-bounded termination checks for larger agent counts. This mirrors the dynamics of real-world scenarios, where misinformation or rumours may not disappear fully from social circles, even if they are countered broadly or corrected.

Another point of interest is the role debunking plays, which was consistently activated early, with full certainty, in all multi-agent models. It is therefore demonstrated that, as soon as the conditions for scepticism are met (for example a debunker neighbour and a susceptible spreader), the correction effect is immediate and deterministic. Regardless, it does not guarantee full correction or stifling of the whole network, suggesting that local effectiveness does not guarantee the collapse of the global system, especially in scale-free networks with persistent spreaders. Therefore, debunking is effective, but insufficient to cause systemic termination on its own. An additional observation is about the stability and correctness of the formal model. All models passed the checks of deadlock detection, and showed the expected growth in state space size, thus validating that the PRISM encoding models the original ABM dynamics. However, the explosion of the state space shows the practical limit of using the exact model checking beyond 6 agents, thus requiring simulations to explore larger models. This demonstrates the scalability limitations of PMC and how important hybrid analysis can be when reproducing larger ABMs. Finally, the combined result confirms and extend the findings presented in the original paper, showing that rumour dynamics remain difficult to extinguish and sensitive to interaction patterns, even when probabilistically formalized. Formal verification offers a complementary perspective to stochastic simulations, as it provides probabilistic guarantees and structural clarity. Moreover, it can also reveal edge cases or divergence points, as well as computational boundaries which can be hidden in simulation traces.

5 DISCUSSION

5.1 Comparison to Existing Literature

This study's results both confirm and extend findings found in previous literature on ABMs and rumour dynamics. Firstly, the trend of diffusion stability increasing and the rumour activating through spontaneous spreading, especially in larger agent populations, is consistent with the observations of Mazzoli et al. (2018) [12]. Their original simulations showed that rumour spread was more likely under low scepticism and moderate influence. Moreover, they explained how network topology plays a significant role in enabling widespread diffusion. In this replication as a formal model, these dynamics have been validated: models with more agents showed certain spontaneous spread and persuasion activation, which suggests that transitions accurately reflected the underlying social mechanisms present in the original ABM. However, unlike the simulations performed in the original paper, this project provides formal certainties and insights into the structure of the model. Considering a methodological perspective, formalizing the ABM as a DTMC and its translation into PRISM aligns with the approaches used by Banisch et al. (2011) [3]. These studies argue that ABMs should be converted into formal probabilistic models in order to enable more rigorous analysis. This work contributes to this concept, demonstrating both the feasibility and limitations of such process when applied to an ABM inspired by the real world. Furthermore, it confirms that formal tools can be used to reproduce and extend ABM simulations, while also revealing the ongoing challenges of scalability and reliability, especially in areas that have high behavioural heterogeneity.

5.2 Evaluation of PRISM as a Tool

During this study, PRISM proved to be a highly valuable platform for formalizing and verifying ABMs, particularly in structured and small or medium scale setting. One of its main strengths is the modular architecture which allows each agent to be defined with its own local state and transition rules. This modularity allowed for fine grained control over how agents behave and facilitated the encoding of the sets of rules. PRISM also supports exact model checking and statistical simulation, thus allowing for flexibility in switching between full behavioural verification and approximate analysis, depending on the size of the model. Another valuable advantage is PRISM's integration of probabilistic temporal logic, more specifically PCTL, which allowed for behavioural queries to be formulated in a readable manner. The ability to use custom labels and reward structures further showcase the expressiveness of this tool, as these features allowed the creation of nuanced properties (like time bounded diffusion or deadlock detection), making PRISM helpful for checking the level of correctness of the model and identifying edge-case outcomes.

However, there were several limitations encountered during the modelling process. A major issue was having to manually duplicate the agent modules. Each additional agent required an almost identical block of code, which soon became inconvenient and prone to errors. Moreover, PRISM lacks the built-in support for generating network, thus requiring the network topology (scale-free structure in this case) to be manually coded. This reduced the flexibility significantly and introduced additional issues to maintain and resolve. Another limitation was the fact that connections between agents are static, since PRISM does not allow state-dependent or dynamically evolving networks. This restricted its use when it comes to modelling real-time or adaptive interaction patterns, which are common in larger ABMs. Finally, debugging guard conditions and resolving conflicts or overlaps from transition logic requires extensive and time-consuming manual checking, especially while working with complex multi-agent interactions. Overall, the PRISM tool is powerful and appropriate for replicating and verifying structured ABMs, that have modest agent counts and fixed rules on interaction. It is mostly well suited for experiments that are controlled and reproducible, and where exhaustive behavioural space coverage is more important than large-scale modelling. PRISM is an excellent choice for studies that prioritize formal guarantees and traceable outcomes; however, for modules which involve dynamic topologies, learning agents or high agent counts, alternative tools or hybrid frameworks (such as formal model verification combined with stochastic simulations) may be more appropriate.

5.3 Implications and Limitations

The results of this study have important insights into intersecting formal methods with social simulations. It shows that ABMS from the social sciences can also be formally verified, rather than solely relying on stochastic (empirical) simulations. This project demonstrates that formally defined behavioural outcomes can be quantified with provable guarantees by translating the rumour-spreading model of Mazzoli et al. (2018) [12] into a format compatible with probabilistic model checking. This work showcases PMC as a way to complement the traditional analysis of ABMs. The findings can also suggest that key dynamics in social models remain robust under format checking, such as debunking consistently occurred within the given timeframe.

However, there are several limitations to be discussed. The greatest constraint was the model size. Since states grew combinatorially, the verified model was restricted to contain at most 10 agents, which limits the discovery of complex and population-wide effects. The network structure was static; thus, it lacked the flexibility of adaptive or rewiring networks. Moreover, agent thresholds had to be manually assigned, in contrast to a more realistic model which could sample them from a distribution or update them dynamically. The model also abstracted certain principles of social behaviours like memory, multi-topic interactions or agent learning, features which are frequently considered in advanced ABMs. Furthermore, some properties of the verification required delicate design choices, for instance avoiding infinite values in reward-based properties due to absorbing paths. These aspects show that, while model checking brings accuracy and rigour, it also necessitates precise specification and effort in encoding, which can become brittle over time. Overall, this study verifies the behaviour of a concrete ABM and maps out a path forward for integrating formal methods into the future of computational social science.

6 CONCLUSION

The aim of this study was to evaluate whether probabilistic model checking (specifically using PRISM) can effectively be used to analyse agent-based models (ABMs) derived from social sciences. More specifically, the model used in this research was proposed by Mazzoli et al. (2018) [12], which evaluated a rumour spreading model. By reproducing and encoding the social sciences ABM, this research demonstrated that it is possible to replicate such models within a formal, small-scaled framework, while also obtaining valuable insights into system behaviours that empirical simulations alone cannot offer. Across 14 properties which were formally verified, the analysis showed high consistency with the simulations presented in the original paper, as well as exposing deeper behavioural structures. The results prove the importance of integrating formal verification into social modelling workflows and simulations.

The analysis shows that the probability of the rumour reaching all agents before it dies out (sub-RQ1) is high across all agents counts, reaching over 99% from 6 agents onwards. However, peers being able to achieve complete influence alone did not occur at all, even within larger populations; the "all influenced" property returned a probability of 0 in all cases. This reveals that full diffusion can occur when spontaneous triggers and persuasion are combined, and not just from influence alone. The system still reached full awareness consistently (non-ignorant states), which supports the assumption that formal modelling is able to capture realistic spreading patterns in decentralized settings.

The verification of the combined trigger property, as well as the final absorption states (sub-RQ2), show that large populations stabilize spread and improve propagation, although the model always starts with a single spreader. Increasing agent counts caused spontaneous and persuasion activation to approach certainty, which indicates that the network size is important for robustness, even when the initial configuration is limited.

For the third sub-RQ, spontaneous spread proved to occur close to 100% of the cases involving more than (or) 3 agents. Persuasion had a steady rise, reaching almost 98% for 10 agents. Debunking was verified to always occur, in both general and within 5 steps, which shows strong correction potential in all of the model configurations. These findings confirm that spontaneous spreading has a critical foundational role with persuasion being a secondary amplifier. Debunking was also proven to be effective but within a localized environment, aligning with the sociological observations from studies on misinformation.

To answer the last sub-RQ, PRISM proved to be sufficient to encode a non-trivial ABM with multiple agents, thresholds and probabilistic transitions, allowing for a rich analysis through its support for reward structures and temporal logic. Regardless, increasing the number of agents becomes a scalability issues, and having to manually duplicate modules limits flexibility. PRISM is henceforth suitable for structured reproducibility studies and verification of modest ABMs. Future research could explore tooling or abstraction extensions in order to expand usability.

Overall, the study demonstrates that probabilistic model checking can successfully verify a wide range of behavioural outcomes in agent-based social simulations. Key mechanisms, such as spontaneous spread, persuasion and debunking were captured accurately through formal properties. The system's emergent behaviour aligned well with the findings from the original paper. PRISM was shown to be an adequate verification tool for such ABMs, when the system is constrained structurally. However, limitations in scalability and dynamic network representation place bounds in how far PMC can be currently extended. Thus, PMC can be used primarily for small- to medium-scale models with well-defined behaviours. It is a valuable complement to stochastic simulation, specifically when strong correctness guarantees are needed.

Future work could extend this researched foundation in multiple directions. A promising front is automating network generation, such that it allows for more realistic social graphs (for example scale-free) to be integrated within PRISM models. Another improvement could be enhancing agent complexity and introducing adaptive or multi-layered behaviours: beliefs, emotions or memory. To overcome the scalability issues, researchers could experiment with abstraction techniques which can reduce state space without losing any analytical power. These directions would enhance the generalisability and impact of formal methods in computational social science.

REFERENCES

[1] Sameera Abar, Georgios K. Theodoropoulos, Pierre Lemarinier, and Gregory M.P. O'Hare. 2017. Agent Based Modelling and Simulation tools: A review of the state-of-art software. *Computer Science Review* 24, (May 2017), 13–33. DOI:https://doi.org/10.1016/j.cosrev.2017.03.001

[2] Robert Axelrod. 2006. Chapter 33 Agent-based Modeling as a Bridge Between Disciplines. *Elsevier eBooks* (January 2006), 1565–1584.
DOI:https://doi.org/10.1016/s1574-0021(05)02033-2

[3] Sven Banisch, Ricardo Lima, and Tanya Araújo. 2012. Agent based models and opinion dynamics as Markov chains. *Social Networks* 34, 4 (October 2012), 549–561. DOI:https://doi.org/10.1016/j.socnet.2012.06.001

[4] Giangiacomo Bravo and Mike Farjam. *Prospects and Challenges for the Computational Social Sciences*. Retrieved April 28, 2025 from https://www.jucs.org/jucs_23_11/prospects_and_challenges_for/jucs_23_11_1057_1069_bravo.pdf

[5] Puneet Chopra, Rasmus Sorensen, Brandon Brown, and Animesh Sharma. 2024. On the limits of agency in agent-based models. *arXiv.org*. Retrieved from https://arxiv.org/abs/2409.10568

[6] Rosaria Conte and Mario Paolucci. 2014. On agent-based modeling and computational social science. *Frontiers in Psychology* 5, (July 2014). DOI:https://doi.org/10.3389/fpsyg.2014.00668

[7] Bruce Edmonds and Scott Moss. 2005. From KISS to KIDS – An "Anti-simplistic" Modelling Approach. *Lecture notes in*

computer science (January 2005), 130–144. DOI:https://doi.org/10.1007/978-3-540-32243-6_11
[8] Joshua M. Epstein. 1999. Agent-based computational models and generative social science. *Complexity* 4, 5 (May 1999), 41–60. DOI:https://doi.org/10.1002/(sici)1099-0526(199905/06)4:5%3C41::aid-cplx9%3E3.0.co;2-f
[9] Christian Hensel, Sebastian Junges, Joost-Pieter Katoen, Tim Quatmann, and Matthias Volk. 2021. The probabilistic model checker Storm. *International Journal on Software Tools for Technology Transfer* (July 2021). DOI:https://doi.org/10.1007/s10009-021-00633-z
[10] Marta Kwiatkowska, Gethin Norman, and David Parker. 2002. PRISM: Probabilistic Symbolic Model Checker. (April 2002), 200–204. DOI:https://doi.org/10.1007/3-540-46029-2_13
[11] Marta Kwiatkowska, Gethin Norman, and David Parker. 2011. PRISM 4.0: Verification of Probabilistic Real-Time Systems. *Lecture Notes in Computer Science* (July 2011), 585–591. DOI:https://doi.org/10.1007/978-3-642-22110-1_47
[12] Mattia Mazzoli, Tullio Re, Roberto Bertilone, Marco Maggiora, and Jacopo Pellegrino. 2018. Agent Based Rumor Spreading in a scale-free network. *arXiv.org*. Retrieved April 28, 2025 from https://arxiv.org/abs/1805.05999
[13] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili. 2007. Theory of rumour spreading in complex social networks. *Physica A: Statistical Mechanics and its Applications* 374, 1 (January 2007), 457–470. DOI:https://doi.org/10.1016/j.physa.2006.07.017
[14] Linhe Zhu, Mengtian Zhou, and Zhengdi Zhang. 2020. Dynamical Analysis and Control Strategies of Rumor Spreading Models in Both Homogeneous and Heterogeneous Networks. *Journal of Nonlinear Science* 30, 6 (May 2020), 2545–2576. DOI:https://doi.org/10.1007/s00332-020-09629-6