Achieving Accurate Opinion Consensus in Large Multi-Agent Systems

by

Oleksandr Pryymak

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

October 2013
UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
Electronics and Computer Science

Doctor of Philosophy

Achieving Accurate Opinion Consensus
in Large Multi-Agent Systems

by Oleksandr Pryymak

Modern communication technologies offer the means to share information within decentralised, large and complex networks of agents. A significant number of examples of peer-to-peer interactions can be found in domains such as sensor networks, social web communities and file-sharing networks. Nevertheless, the development of decentralised systems still presents new challenges for sharing uncertain and conflicting information in large communities of agents. In particular, the problem of forming opinion consensus supported by most of the observations distributed in a large system, is still challenging.

To date, this problem has been approached from two perspectives: (i) on a system-level, by analysing the complex processes of opinion sharing in order to determine which system parameters result in higher performance; and (ii) from the perspective of individual agents, by designing algorithms for interactively reaching agreements on the correct opinion or for reasoning about the accuracy of a received opinion by its additional annotation.

However, both of these approaches have significant weaknesses. The first requires centralised control and perfect knowledge about the configuration of the system in order to simulate it, which are unlikely to be available for large decentralised systems. Whereas, the latter algorithms introduce a significant communication overhead, whilst in many cases the capabilities of the agents are restricted and communication strictly limited. Therefore, there is a need to fill the gap between these two approaches by addressing the problem of improving the accuracy of consensus in a decentralised fashion with minimal communication expenses.

With this motivation, in this thesis we focus on the problem of improving the accuracy of consensus in large, complex networks of agents. We consider challenging settings in which communication is strictly limited to the sharing of opinions, which are subjective statements about the correct state of the subject of common interest. These opinions are dynamically introduced by a small number of sensing agents which have low accuracy, and thus the correct opinion just slightly prevails in the readings. In order to form the
accurate consensus, the agents have to aggregate opinions from a number of sensing agents which, however, they are very rarely in direct connection with.

Against this background, we focus on improving the accuracy of consensus and develop a solution for decentralised opinion aggregation. We build our work on recent research which suggests that large networked systems exhibit a mode of collective behaviour in which the accuracy is improved. We extend this research and offer a novel opinion sharing model, which is the first to quantify the impact of collective behaviour on the accuracy of consensus. By investigating the properties of our model, we show that within a narrow range of parameters the accuracy of consensus is significantly improved in comparison to the accuracy of a single sensing agent. However, we show that such critical parameters cannot be predicted since they are highly dependent on the system configuration.

To address this problem, we develop the Autonomous Adaptive Tuning (AAT) algorithm, which controls the parameters of each agent individually and gradually tunes the system into the critical mode of collective behaviour. AAT is the first decentralised algorithm which improves accuracy in settings where communication is strictly limited to opinion sharing. As a result of applying AAT, 80-90% of the agents in a large system form the correct opinion, in contrast to 60-75% for the state-of-the-art message-passing algorithm proposed for these settings, known as DACOR. Additionally, we test other research requirements by evaluating teams with different sizes and network topologies, and thereby demonstrate that AAT is both scalable and adaptive. Finally, we showed that AAT is highly robust since it significantly improves the accuracy of consensus even when only being deployed in 10% of the agents in a large heterogeneous system.

However, AAT is designed for settings in which agents do not differentiate their opinion sources, whilst in many other opinion sharing scenarios agents can learn who their sources are. Therefore, we design the Individual Weights Tuning (IWT) algorithm, which can benefit from such additional information. IWT is the first behavioural algorithm that differentiates between the peers of an agent in solving the problem of improving the accuracy of consensus. Agents running IWT attribute higher weights to opinions from peers which deliver the most surprising opinions. Crucially, by incorporating information about the source of an opinion, IWT outperforms AAT for systems with dense communication networks. Considering that IWT has higher computational cost than AAT, we conclude that IWT is more beneficial to use in dense networks while AAT delivers a similar level of accuracy improvement in sparse networks, but with a lower computational cost.
# Contents

Nomenclature .................................................. xiii

Acknowledgements .............................................. xvii

Declaration of Authorship ..................................... xix

1 Introduction ................................................. 1
   1.1 Complexity in Large Multi-Agent Systems .................. 3
   1.2 Opinion Formation ........................................ 5
   1.3 Research Aims ........................................... 7
   1.4 Research Requirements .................................. 9
   1.5 Research Contributions ................................ 11
   1.6 Thesis Structure ....................................... 13

2 Related Work ............................................... 15
   2.1 Consensus in Large Systems ............................. 15
       2.1.1 Accuracy of Consensus .............................. 17
       2.1.2 Consensus in Decentralised Systems .............. 18
   2.2 Modelling Large Multi-Agent Systems .................. 19
       2.2.1 Structural Properties of Decentralised Systems 21
       2.2.2 Networks of Interactions .......................... 24
       2.2.3 Dynamic Processes on Networks .................. 30
   2.3 Models of Opinion Sharing .............................. 33
       2.3.1 Models with Dynamically Introduced Opinions 35
       2.3.2 Opinion Cascades .................................. 39
   2.4 Benchmark Algorithm for Achieving Accurate Consensus 42
   2.5 Summary ................................................ 44

3 Modelling Collective Behaviour in Opinion Sharing .......... 47
   3.1 Opinion Sharing Model .................................. 48
       3.1.1 Model Description .................................. 48
       3.1.2 Performance Metrics ................................ 54
       3.1.3 Comparison Against the Existing Models .......... 59
   3.2 Computational Modelling Methodology .................... 63
       3.2.1 Goals of the Computational Modelling ............. 64
       3.2.2 Experimental Setups ............................... 65
       3.2.3 Simulation Process ................................ 70
   3.3 Social Dynamics in the Model ........................... 71
3.3.1 Critical Mode with the Highest Accuracy of Consensus . . . . . . . 73
3.3.2 Opinion Dynamics . . . . . . . . . . . . . . . . . . . . . . . . . . . 76
3.4 Analysis of the Critical Mode . . . . . . . . . . . . . . . . . . . . . . . . 78
  3.4.1 Accuracy Improvement . . . . . . . . . . . . . . . . . . . . . . . . . 79
  3.4.2 Critical Agents’ Weights . . . . . . . . . . . . . . . . . . . . . . . . 79
  3.4.3 Communication Expense . . . . . . . . . . . . . . . . . . . . . . . . 80
  3.4.4 Convergence to Consensus . . . . . . . . . . . . . . . . . . . . . . . 81
  3.4.5 Branching Factor as Indicator on the Critical Mode . . . . . . . . 82
3.4.6 Influence of the System Size . . . . . . . . . . . . . . . . . . . . . . . 82
3.5 Centralised Selection of the Critical Agents’ Weight . . . . . . . . . . . 84
  3.5.1 When Information about the System is Perfect . . . . . . . . . . . . 85
  3.5.2 When the System Parameters are Known . . . . . . . . . . . . . . . 86
  3.5.3 When the System Parameters are Unknown . . . . . . . . . . . . . . 87
  3.5.4 Empirical Evaluation . . . . . . . . . . . . . . . . . . . . . . . . . . 87
3.6 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87

4 Accurate Consensus with Anonymous Peers 91
  4.1 The Autonomous Adaptive Tuning Algorithm . . . . . . . . . . . . . . . 92
    4.1.1 Awareness Rate as an Indicator of the Critical Mode . . . . . . . . 93
    4.1.2 Candidate Weights . . . . . . . . . . . . . . . . . . . . . . . . . . . 96
    4.1.3 Dynamic Estimation of the Awareness Rates . . . . . . . . . . . . . 97
    4.1.4 Weight Selection Strategies . . . . . . . . . . . . . . . . . . . . . . 98
  4.2 Reducing Computational Cost of AAT Algorithm . . . . . . . . . . . . . . 99
    4.2.1 Limiting the Set of Candidate Weights . . . . . . . . . . . . . . . . 100
    4.2.2 Heuristic Estimation of the Awareness Rates . . . . . . . . . . . . . 103
  4.3 Analysis of AAT Parameters . . . . . . . . . . . . . . . . . . . . . . . . . 105
    4.3.1 Selection of the Target Awareness Rate . . . . . . . . . . . . . . . . 107
    4.3.2 Comparison of AAT Weight Selection Strategies . . . . . . . . . . . 107
  4.4 Empirical Evaluation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 109
    4.4.1 Accuracy of Consensus . . . . . . . . . . . . . . . . . . . . . . . . . 110
    4.4.2 Communication Expense . . . . . . . . . . . . . . . . . . . . . . . . 112
    4.4.3 Computational Expense . . . . . . . . . . . . . . . . . . . . . . . . 113
    4.4.4 Heterogeneous Agent Population . . . . . . . . . . . . . . . . . . . 115
  4.5 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 117

5 Accurate Consensus with Identified Peers 121
  5.1 The Problem of Determining Preferences among Peers . . . . . . . . . . . 122
    5.1.1 The Number of Variable Weights in the System . . . . . . . . . . . . 122
    5.1.2 Opinion Sharing in Dense Communication Networks . . . . . . . . . 123
    5.1.3 Agents’ Local Views . . . . . . . . . . . . . . . . . . . . . . . . . . 124
    5.1.4 Behavioural Algorithms for Determining Preferences among Peers 125
  5.2 Limiting Connectivity According to the Preferences . . . . . . . . . . . . 126
    5.2.1 Randomised Preferences as a Benchmark . . . . . . . . . . . . . . . 127
    5.2.2 Preferences by the Shortest Path Length to Opinion Sources . . . . 128
    5.2.3 Preferences by Opinion Timeliness . . . . . . . . . . . . . . . . . . 130
    5.2.4 Preferences by Opinion Surprise . . . . . . . . . . . . . . . . . . . 133
  5.3 The Individual Weight Tuning Algorithm . . . . . . . . . . . . . . . . . . 136
CONTENTS

5.4 Empirical Evaluation ........................................... 139
   5.4.1 The Accuracy of Consensus .......................... 139
   5.4.2 Communication Expense ............................... 141
   5.4.3 Computational Expense ................................. 142
5.5 Summary ....................................................... 144

6 Conclusions .................................................... 147
   6.1 Conclusions ................................................ 147
   6.2 Future Work ............................................... 151

A Additional Results for the Model Evaluation .................. 155

B Additional Results for AAT Evaluation ....................... 157

C Additional Results of IWT Evaluation ....................... 161

References ...................................................... 163
## List of Figures

1.1 Position of our research in regards to other fields .......................... 8  
2.1 Example of a business-oriented social network ............................... 20  
2.2 The clustering coefficient as a measure of the interconnectivity .......... 24  
2.3 Comparison of Poisson and power-law degree distribution .................. 26  
2.4 Small-world topology generator .............................................. 28  
2.5 Sample system in the process of opinion sharing ............................ 38  
2.6 Number of belief updates $n_o(w_i)$ in Glinton et al.’s model ............ 40  
2.7 Different cases of dynamics in Glinton et al.’s model ..................... 42  
2.9 The size of opinion cascade depending on the average belief .............. 43  
3.1 Sample structure of a multi-agent system .................................. 50  
3.2 Model of an agent .................................................................... 51  
3.3 Agent’s aggregation functions .................................................. 52  
3.4 Agent’s decision rule ............................................................... 53  
3.5 Sample dynamics of the model .................................................. 54  
3.6 Star topology ........................................................................... 55  
3.7 Theoretical bounds on the accuracy of consensus ............................ 57  
3.8 Accuracy of consensus in ours and Glinton et al.’s models ............... 61  
3.9 Sources of the accuracy improvement in our and Glinton et al.’s models 62  
3.10 Model performance and the distribution of priors of agents’ beliefs ... 67  
3.11 Average shortest path length of the communication networks .......... 70  
3.12 Clustering coefficient of the communication networks .................... 70  
3.13 Accuracy of consensus depending on the common weight ............... 73  
3.14 Communication expenses depending on the common weight ............ 75  
3.15 Convergence to consensus depending on the common weight .......... 76  
3.16 Branching factor depending on the common weight ...................... 77  
3.17 Variance in opinion cascade sizes depending on the common weight .... 78  
3.18 Highest accuracy achieved in the critical mode ............................ 79  
3.19 Agents’ weights in the critical mode ........................................... 80  
3.20 Communication expenses in the critical mode .............................. 81  
3.21 Convergence to consensus in the critical mode ............................. 82  
3.22 Branching factor in the critical mode ........................................... 83  
3.23 Highest accuracy in systems of different sizes ............................. 84  
3.24 Critical weights in systems of different sizes ............................... 84  
3.25 Convergence to consensus of the systems of different sizes .......... 85  
3.26 Accuracy of consensus achieved by the benchmarks ..................... 88
4.1 Activity diagram of an agent executing AAT ........................................ 94
4.2 Awareness rate as an indicator of the critical mode .............................. 95
4.3 Sample dynamics of an agent’s belief ................................................ 101
4.4 Selecting the target awareness rate $h_{trg}$ for AAT ............................ 106
4.5 Selecting the strategy for AAT ............................................................. 108
4.6 Sample dynamics of AAT with different strategies .............................. 109
4.7 Accuracy of consensus achieved by AAT and DACOR ....................... 112
4.8 Communication expenses for AAT and DACOR ................................. 114
4.9 Computational expenses for AAT and DACOR as the number of weights changes ................................................................. 115
4.10 Computational expenses for AAT and DACOR ................................... 116
4.11 Accuracy of consensus for AAT and DACOR in heterogeneous system ... 117
5.1 The accuracy of consensus with randomised preferences ..................... 128
5.2 Sample weights selected by the shortest path length to any sensing agent 129
5.3 The accuracy of consensus with preferences as the shortest path length to opinion sources ................................................................. 130
5.4 Sample weights selected by the timeliness of received opinions .......... 132
5.5 The accuracy of consensus with preferences as timeliness of received opinions 132
5.6 Sample weights selected by the opinion surprise of the peers .......... 135
5.7 The accuracy of consensus with preferences as opinion surprise .......... 135
5.8 Sample weights selected by the IWT algorithm .................................. 137
5.9 Activity diagram of the agent executing IWT .................................... 138
5.10 The accuracy of consensus achieved by IWT .................................... 140
5.11 Communication expenses for IWT ...................................................... 142
5.12 Computational expenses for IWT ....................................................... 143
A.1 Highest accuracy and agents weights in the critical mode ................... 156
A.2 Branching factor as an indicator of the critical mode of behaviour ........ 156
B.1 Accuracy of consensus achieved by AAT and DACOR with the Bayesian aggregation function on dense network ........................................ 158
B.2 Accuracy of consensus achieved by AAT with the Weighted Sum aggregation function ................................................................. 159
C.1 Accuracy of consensus achieved by IWT with the Weighted Sum aggregation function ................................................................. 162
List of Tables

3.1 Experimental setups for the model evaluation ........................................ 71
4.1 Experimental setups for the algorithms evaluation .................................... 110
5.1 Complexity of the search space of the algorithms ................................... 123
5.2 Local view of an agent ............................................................................. 125
List of Algorithms

1  DACOR: Benchmark Algorithm for Achieving Accurate Consensus . . . . 44
2  AAT: Hill-climbing strategy to select a weight . . . . . . . . . . . . . . . . 99
3  AAT: Generation of the Candidate Weights . . . . . . . . . . . . . . . . 103
4  iAAT: Awareness Estimation Rules . . . . . . . . . . . . . . . . . . . . . . 105
5  AAT: Limiting the agent’s connectivity following its preferences . . . . . 127
Nomenclature

\(N\) \hspace{1cm} \text{number of agents in the multi-agent system (150...10000)}

\(N_s\) \hspace{1cm} \text{number of sensing agents (5\% of } N)\)

\(A\) \hspace{1cm} \text{set of agents}

\(S\) \hspace{1cm} \text{set of sensing agents}

\(\langle d \rangle\) \hspace{1cm} \text{average number of neighbours (the expected network degree, 6...100)}

\(M\) \hspace{1cm} \text{set of independent opinion sharing rounds (50 or 500 rounds)}

\(B\) \hspace{1cm} \text{set of possible states of the subject of common interest (2 alternatives)}

\(b\) \hspace{1cm} \text{correct opinion on the current opinion sharing round}

\(\text{the true state of the subject of common interest)}\)

\(p_i^t\) \hspace{1cm} \text{prior belief of agent } i \ (0...1)

\(p_i^k\) \hspace{1cm} \text{belief of agent } i \text{ on update step } k \ (0...1)

\(o_i^k\) \hspace{1cm} \text{opinion of agent } i \text{ on update step } k \ (B \cup \{\text{undetermined}\})

\(D_i\) \hspace{1cm} \text{set of agents in network neighbourhood of agent } i \ (E[|D_i|] = \langle d \rangle)

\(w_{ij}\) \hspace{1cm} \text{weight agent } i \text{ assigns to opinion of its neighbour agent } j \ (0.5...1)

\(\{\sigma, 1 - \sigma\}\) \hspace{1cm} \text{confidence bounds of the agent’s belief,}

\(\text{upon crossing which the agent changes its opinion } \{\{0.2, 0.8\}\})\)

\(r\) \hspace{1cm} \text{accuracy of new opinions (65\%)}

\(R\) \hspace{1cm} \text{accuracy of consensus (3.25...100\%)}

\(R_{\text{min1}}\) \hspace{1cm} \text{accuracy of consensus}

\(\text{when opinions are not shared in the system (3.25\%)}\)

\(R_{\text{min2}}\) \hspace{1cm} \text{accuracy of consensus}

\(\text{when a single new opinion shared on the system scale, (63.86\%)}\)

\(R_{\text{max}}\) \hspace{1cm} \text{theoretical maximum of the accuracy of consensus (79...99\%)}

\(U\) \hspace{1cm} \text{communication as the number of messages transmitted in the system}

\(U_{\text{min}}\) \hspace{1cm} \text{minimum number of messages required to share}

\(\text{a single opinion to all agents}\)

\(C\) \hspace{1cm} \text{convergence as the number of update steps before reaching a consensus}

\(C_{\text{min}}\) \hspace{1cm} \text{minimum number of update steps required to shared}

\(\text{a single opinion to all agents}\)

\(\lambda\) \hspace{1cm} \text{rate of opinions introduction (10 update steps between observations)}

\(\Lambda\) \hspace{1cm} \text{total number of introduced opinions (3 } \cdot \ N_s\)
Acknowledgements

I would like to thank all of the people who encouraged and supported me during the undertaking of my research.

First and foremost, I would like to sincerely thank my supervisors Prof. Nick Jennings and Prof. Alex Rogers to whom I am greatly indebted for their support, guidance and encouragement. I am incredibly thankful the opportunity they have provided me to do this research, and for their expertise in this fascinating field.

I also thank the ALADDIN project\textsuperscript{1} and the School of Electronics and Computer Science at University of Southampton for funding the completion of this PhD.

Furthermore, I would like to thank all of the members of the Agents, Interaction and Complexity group at the University of Southampton who were of great support throughout my PhD, particularly Ruben Strangers, Zinovi Rabinovich, Enrique Munoz De Cote, Muddasser Alam, Colin Williams, Harry Rose, Sebastian Stein, Long Tran-Thanh, Oliver Parson and Matteo Venanzi. It has been a pleasure to work and socialise with such a great team, both on a day-to-day basis and at the conferences we have attended together.

Finally, I thank my parents and my sister for their support, who have always impressed upon me the importance of education and have always supported me through the highs and lows of academic life.

\textsuperscript{1}funded jointly by BAE Systems and Engineering and Physical Sciences Research Council (EPSRC), grant reference EP/C548051/1
Declaration of Authorship

I, Oleksandr Pryymak, declare that the thesis entitled

Achieving Accurate Opinion Consensus
in Large Multi-Agent Systems

and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

• this work was done wholly or mainly while in candidature for a research degree at this University;
• where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
• where I have consulted the published work of others, this is always clearly attributed;
• where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
• I have acknowledged all main sources of help;
• where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
• parts of this work have been published as:

Signed: .................................................................

Date: .................................................................

xix
Chapter 1

Introduction

We live in a data-rich world. However, much of the most important information is highly dispersed and we can make a use of it only by being equipped with a smart process of aggregation and filtering. This dispersal was caused by technological development and the shift from centralised to decentralised information systems, as the scale of the latter increased. This has deeply affected our society and generated new technological challenges. For example, previously only major news agencies were able to aggregate a large number of reports in a centralised fashion, to form an accurate, but still subjective, opinion about the ongoing events, and inform all their subscribers. By contrast, we currently face the proliferation of a new type of media — social media that is based on peer-to-peer interactions, such as Facebook, Twitter and the Ushahidi platform\footnote{A crowd-sourcing platform for social activists which gathers citizen reports in order to create a temporal and geospatial archive of events, \url{http://ushahidi.com/}}. In such platforms individuals have become much closer to the witnesses of events and their opinions, but now they are facing a difficult task of forming their own accurate opinion, which correctly corresponds to reality. Unlike in centralised systems, there is no central authority which aggregates all the potentially conflicting opinions of witnesses and then develops its own assessment, which is likely to be the most accurate opinion. Instead, in these networked societies, which are shaped by relations between participants, each individual forms their own opinion under the influence of their peers. This individual then in turn re-influences its peers with its new opinion, and thus spreads it further, participating in a cascade of opinion sharing. Usually, this results in a number of groups supporting different opinions, and therefore achieving an accurate opinion consensus becomes a challenging problem.

This problem, which is essentially a distributed information aggregation problem, is not limited to human societies and is a crucial aspect of many decentralised systems. Similarly to media agencies, centralised computer mainframes could not cover all our needs. As the result, a vast number of decentralised information systems were developed, such as sensor networks for distributed monitoring, networks of mobile devices and the
Internet. Despite distinct areas, these information systems, online societies and even a number of biological systems share common properties of complex relations between their participants, such as hardware or software agents, people, and even bees in a swarm. In these systems a large number of autonomous participants are constantly interacting, sharing information and forming their own state under the influence of others. To give examples of such systems, which face the problem of reaching an accurate consensus in a decentralised manner, we offer the following scenarios:

1. Imagine we are given a task to automate monitoring of unfolding disasters, like earthquakes that happened in Chile or Haiti in 2010 or the political and subsequently humanitarian crisis in Egypt, 2011. In these cases, the online social networks, such as Twitter or the Ushahidi platform, provide a vast amount of citizen reports. However, two key issues arise here: (i) How to form an accurate opinion about the events relying only on a limited number of peers, which may report conflicting information? (ii) How to facilitate the emergence of an accurate consensus? To date, we do not have techniques that can provide complete answers to these questions.

In order to approach these questions and develop a corresponding personal software agent to assist us, or a number of autonomous decision-making agents to help a whole society, we have to consider that online societies are large and consist of thousands and even millions of individuals. At the same time, only a few of them are actually located at the scene of event and they often make quite uncertain and misleading observations. However, all of the observations cannot be communicated to every individual in the society, because on a large scale this would have lead to an information overload, while the individuals usually have limited resources (Toffler, 1970). As a result, individuals cannot aggregate all existing observations in order to form an accurate opinion. To mitigate overload, each individual has to filter all incoming information and only communicate what is useful to its peers (Shapiro and Varian, 1999, chap.1). Therefore, communication is usually limited to the exchange of conclusions that are opinions about the ongoing events. Each *opinion* is a subjective statement about the correct state of a common subject of interest that is shared without any supporting information that led to its formation. Considering these restrictions the questions we pose become more challenging. To answer them, we have to clarify under which conditions the individuals are able to reach an accurate consensus, and how to design them in order to elicit such behaviour in any society.

2. Now, imagine that a system of thousands of microscopic sensors is deployed to monitor the city of London, or even the surface of another planet. To cover such a vast scale and reduce their cost, these sensors have to be very efficient, and thus quite limited in their capabilities. Hence, they have to limit a number of
their interactions to a relatively small neighbourhood of peers, and minimise communication in order to conserve their battery charge and the bandwidth of the communication channel.

Despite being an artificial system, its properties, such as the peer-to-peer topology of interactions and the limitation in communication, are very similar to the previous scenario. For example, such a distributed sensor network might need to reach a consensus on operational issues, such as to choose the least noisy communication channel or to make a decision on switching from sleep to a fully-operational mode, given only a few noisy observations dispersed in the system. So, how can these microscopic sensors efficiently reach an accurate consensus and benefit from their large number?

Our research addresses the questions we posed above. To abstract from a specific application, we analyse these problems from the perspective of a multi-agent systems paradigm (Jennings, 2001). The multi-agent paradigm provides a suitable description of such systems, and amongst others, it was successfully applied to the analysis of opinion sharing in large societies (Castellano et al., 2009) and the investigation of the concept of emergence (Serugendo et al., 2006). These are essential steps towards our aim of designing a decentralised solution to foster the emergence of an accurate consensus. Crucially, to develop an efficient and scalable solution that can be applied to diverse scenarios, we assume that communication is limited and the agents are able to share only their opinions. By tackling the problem under such restriction, we aim to turn the complexity introduced by the size of a large multi-agent system from its weakness into its power.

### 1.1 Complexity in Large Multi-Agent Systems

Large networked multi-agent systems, as described above, exhibit an enormous increase in complexity in comparison to traditional centralised information systems. This is due to their decentralised and distributed nature, where potentially millions of heterogeneous and dynamic agents interact. Since the number of possible interactions is combinatorial in the number of agents, this poses new challenges as traditional engineering approaches are often inadequate to address the dynamism and uncertainty that are inherent in such systems (Raje and Chinnasamy, 1999). To address these challenges we need to analyse processes in existing large multi-agent systems, such as our society or biological communities, and draw insights from physical systems that share similar properties.

Crucially, however, despite their size, these large multi-agent systems often demonstrate cohesion as a collective result of individual action. Systems such as a flock of birds choosing a direction of flight, processes of rumours and infections spreading in a society, and even cascades in a growing pile of sand are difficult to predict, but they still have
characteristic statistical properties (Ball, 2012). The field that tackles these problems and describes systems of this sort is often called \textit{complexity science} (Mitchell, 2009). A number of different definitions agree on the general consensus that a \textit{complex system} is one made from a number of components (or agents, if these components represent entities that can make decisions) that interact strongly with one another and therefore, as a result of this, its behaviour cannot be explained in terms of its individual components (Anderson, 1972). One of the first who pointed to the fact that society is a complex system was Schelling (1978) with his book “Micromotives and Macrobphere”. He showed that decision-making is an interactive social process and its outcome is not always predictable from an inspection of individual behaviour.

The key observation drawn from theoretical analysis and supported by simulations is that complexity of interaction does not necessarily lead to chaos and unpredictability. Specifically, societies are often characterized by stunning global regularities despite their large number of participants (Ball, 2005). These modes of organised \textit{collective behaviour}, such as the coherent motion of a bird flock or the formation of an opinion consensus in a society, emerge from a vast number of individual interactions. This ability of complex systems to demonstrate patterns of ordered behaviour that arise from the bottom rather than being imposed by an authority is often called \textit{self-organisation} (Serugendo et al., 2006). Typically, a self-organised mode appears suddenly after a small change and results in the global change of a system state. For example, sudden change in physical substances, such as from a frozen phase to a liquid phase, is a collective property that depends on the interaction between molecules. Similar to these processes, changes in models of social behaviour often exhibit comparable \textit{phase transitions} (Hoyst et al., 2000; Levy, 2005). Another common characteristic of complex systems is the presence of fluctuations and variations of many sorts and scales. For example, a growing pile of sand settles into a static state when no grains are added. However, if grains are continually dropped from above, energy accumulates in the growing steep slopes and just a few grains may disrupt the whole slope releasing a \textit{cascade}. In complex social systems cascades are very common and can be observed in epidemics of contagious disease, or spreading panic, fads and rumours (Bikhchandani et al., 1992; Easley and Kleinberg, 2010), as well as in opinion formation processes (Watts, 2002; Glinton et al., 2009), which is highly relevant to the topic of interest. Fluctuations in the system state, that are caused by cascading behaviour, are dependent on the specific properties of the society, and thus, the behaviour of complex systems is hard to predict, and even harder to control. The only reliable approach to influencing such societies, and changing their behaviour in a decentralised manner, is to facilitate the emergence of a self-organised mode which introduces the properties desired. In the context of our problem, we focus on discovering and exploiting self-organised modes that can lead to the formation of an accurate consensus.
1.2 Opinion Formation

In order to approach our problem in such complex settings we can draw insight from existing studies of opinion sharing processes. Recently, with the massive popularisation of online social networks, this field become one of the most important areas in social studies (Ball, 2012). Specifically, this development has attracted new research into mechanisms and dynamics of opinion sharing. For example, in 2010 after an earthquake in Chile researchers investigated how opinions on Twitter can be trusted in such a disaster response scenario, and they showed that correct opinions exhibit distinct sharing dynamics (Mendoza et al., 2010). In the same manner, records on social networks were scrutinised during the Arab Spring of 2011 (Beaumont, 2011) and the spread of rioting and looting across the United Kingdom in response to a seemingly irrelevant local outbreak of violence in 2011 (Gross, 2011; Hari, 2011). These events showed how, under certain conditions, small changes can cause opinions to cascade on a large scale, and the data collected by researchers enabled them to validate a number of the opinion sharing models that has been proposed in the literature over the last two decades.

In more detail, models of such complex systems mainly use simplistic schemes to describe micro-processes of social influence and are mainly focused on analysing emergent macro-level behaviour (Castellano et al., 2009). The early models of opinion sharing are very similar to statistical physics approaches, where statistical methods are applied to explain how the interaction of a large number of components may exhibit behaviour patterns. In these models, the agents update their internal state through randomised interaction with their peers and the emergent macroscopic behaviour of the system is the aggregate of all these interactions. In particular, opinion formation models such as the Voter model (Krapivsky, 1992), the Sznajd model (Sznajd-Weron, 2000) and the majority rule model (Galam, 2002) share similar properties with physical models of magnetism such as the Ising model (Binney et al., 1992; Young, 2006). Such simplified and abstract opinion sharing models are well understood now from the perspective of formal physics, however, due to their simplicity they do not accurately describe processes that happen in a real society. Critically, as observed during recent events we mentioned above, these models are too abstract to draw conclusions about the real world.

In order to extend such theoretical findings, studies of opinion sharing now commonly try to inject more real-world relevance by relying on computational, agent-based models (Wooldridge, 2002; Mesbahi and Egerstedt, 2010). This modelling approach enables us to incorporate a number of crucial properties. Specifically, the complex topological structures of a network of interactions, which makes a big difference to collective behaviour. Randomised opinion sharing, as used in early models, does not reflect the underlying interaction network, its social ties and degrees of social influence. The process of introducing new opinions into the system also plays a crucial role, since these
changes contribute to modelling cascading behaviour. Finally, the agents’ decision making processes were extremely simplistic in the early models. Crucially, by studying the dynamics of models which incorporate such features, Bikhchandani et al. (1992) demonstrated that opinion sharing occurs in the form of opinion cascades (or “avalanches”), which are a characteristic sign of a complex behaviour. This shows how a single new observation may trigger a large number of agents to alter their opinions and cause a sudden change in the system state. Subsequently, it was shown that such systems exhibit complex emergent behaviour in sharing processes (Watts, 2002) that, in some cases, can be exploited.

Specifically, a model offered by Glinton et al. (2009, 2010a) suggests that collective behaviour influences the accuracy of shared opinions. In this model, the agents aggregate the opinions of their peers with a certain weight, which encodes the number of opinions that an agent has to receive in order to adopt the same opinion and propagate it further. Clearly, this weight is the key factor in influencing the dynamics of the opinion sharing process, and it was found that within a particular range of weights, the accuracy of consensus significantly improves compared to the accuracy of the opinions introduced into the system. Their analysis showed that this state corresponds to a specific critical mode of collective behaviour which is characterised by a power-law distribution of the sizes of opinion cascades. This critical mode of collective behaviour implements a distributed opinion aggregation on a scale of the system. Frequent smaller cascades prevent the multi-agent system from overreacting to early and possibly inaccurate opinions, and only a few large cascades occur to disseminate locally supported opinions to the rest of the agents. Such critical mode of behaviour corresponds to a phase transition between a stable mode of the opinion sharing process (when opinions are not shared) and an unstable mode (when the first opinion, which is possibly incorrect, is shared on a system-wide scale). However, the range of weights which induce this collective behaviour is very narrow and very sensitive to the configuration of the system. This finding suggests that collective behaviour can be exploited in order to improve the accuracy of consensus and is a promising step towards solving our problem.

However, in this model, new observations are introduced in such a way that agents locally filter them before forming and sharing their own opinions. Unfortunately, this implies that (i) agents with sensors may never form and share their opinions if they do not receive enough observations; (ii) speed of convergence to the consensus cannot be measured, since sharing of the observations is delayed due to the local filtering; (iii) improvement of the accuracy of consensus is a combination of collective behaviour and a particular design of the local filtering procedure. Crucially, the latter implies that the specific impact of collective behaviour on the accuracy of consensus is not clear. Thus, there is a need to address these existing shortcomings before approaching our problem.

Despite this, in terms of their model, researchers have successfully addressed the problem of self-organising agents in the described critical mode and presented the Distributed
Adaptive Communication for Overall Reliability (DACOR) algorithm (Glinton et al., 2010a). However, DACOR is a message passing algorithm which requires significant communication overhead in order for agents to exchange new control messages. Furthermore, large systems are often heterogeneous and its agents would not be able to extend their protocol simultaneously to all its agents. Additionally, as our empirical evaluation reveals, the internal parameters of DACOR are sensitive to the configuration of the system and they have to be tuned individually for different domains. Clearly, due to these shortcomings we cannot apply such a solution to the large systems discussed above. It is against this background that we now define own research aims.

1.3 Research Aims

In our motivating scenarios each agent has an objective of forming its own opinion about the subject of common interest. Crucially, this has to be the correct opinion, which reflects the true state of the subject of interest and is expected to be supported by the majority of the observations. When agents objectives are combined, this leads to an overall objective of reaching the accurate consensus. Formally, the accuracy of consensus is the probability of forming the correct opinion by each agent. Considering our discussion that reaching consensus in a decentralised manner is a challenging task, our main research aim is to develop a solution which assists in this, and crucially, improves the accuracy of consensus. The latter can be achieved by designing efficient methods for aggregating available observations, which might be highly distributed in large systems.

The challenge is to solve this problem in settings where the agents’ communication is minimal by restricting agents to only sharing opinions without any supporting information that may assist in opinion aggregation. Following our motivating scenarios, this restriction can be found in many real-world systems where: (i) communication is limited or expensive (e.g. distributed sensor networks); (ii) communication cannot be extended (e.g. in large heterogeneous systems with established communication protocols); and (iii) in human-agent societies where the behaviour of some participants cannot be altered, or agents do not have enough resources or skills to analyse the original information.

This problem of aggregating noisy observations scattered in a system to reach an accurate consensus has attracted a large amount of interest in a number of research communities. However, most of the solutions offered require additional communication to operate and thus are unacceptable for our settings, e.g. agreement protocols which require a large number of interactions to converge (Olfati-Saber et al., 2007) or reasoning about the accuracy of communicated information by its annotation (Moreau, 2009). Given restricted communication, as defined above, the aim of each agent is to filter out incorrect opinions in the process of their aggregation and thereby form the correct one relying
only on their peers. The computational trust community developed a number of similar models in which agents form their beliefs using weighted aggregations of received information, with the weights defined by the trustworthiness of their peers (Ramchurn et al., 2004). However, in such models agents learn these weights based on an eventual observation of the subject’s correct state. This assumption is unlikely to be held in large distributed systems, which are missing a centralised authority which can make an accurate aggregation of all the available opinions. On other side, distributed trust models developed for pure ad-hoc networks do not suffer from this weakness (Pirzada and Mcdonald, 2006). However, similarly to many service-oriented models, in ad-hoc trust models the trustworthiness of a peer is defined as a rate of successful fulfilment of some requests. This approach cannot be extended to information gathering scenarios. Specifically, similarly to the centralised trust models, success in making an accurate observation can be measured only in comparison to the subject’s correct state which is not directly observable.

Against this background, in this work we focus on the development of an efficient solution for improving consensus accuracy within the constraints of minimal communication. Figure 1.1 illustrates how our research aims are positioned relative to existing approaches. Optimal solutions, such as agreement protocols and algorithms that annotate communicated information, require excessive communication overhead which grows with the number of agents and that results in an upper limit of system size. On the other side, modelling of opinion formation generally focuses on very large systems in order to

Figure 1.1: Position of our research in fields studying opinion formation and methods for improving the accuracy of consensus.
minimise the influence of specific individuals and other systemic irregularities. The models we briefly discussed above only provide certain insights into dynamic processes and emerging collective behaviour but do not, however, lead to applied recommendations for improving real-world societies. Nonetheless, the insights gained from modelling opinion formation helped to develop methods that identify influential agents within a society (Bakshy et al., 2011); predict the outcome of the opinion sharing process (Kimura et al., 2010); suggest how the topology of a communication network should be altered in order to improve sharing processes (Watts, 2003); and many others.

However, little attention has been given to the development of self-organisation techniques that will elicit desired behaviours. Such decentralised algorithms will rely on the properties of collective behaviour, and thus, will not be able to reach an optimal solution due to the fluctuations that are often present in complex systems. Crucially, unlike the optimal algorithms, such an approach might not have an upper limit of system size. Since we intend to exploit the properties of collective behaviour that are known to be very sensitive to system properties, our approach should satisfy a number of requirements. Specifically, it has to be independent from the nature of shared opinions and be easily extended to the discussed networked societies. In the following section we discuss such specific requirements in detail.

1.4 Research Requirements

To meet our research aims we need to develop a decentralised solution that self-organises a multi-agent system into a mode in which the accuracy of consensus formation is higher. However, the developed solution has to deal with a number of issues such as large numbers of agents and significant resource constraints. Given this, we identify the following broad research requirements that the solution has to satisfy:

**Requirement 1: Improving the Accuracy of Consensus**

Following the aims of our research, the main requirement is to reach a high level of *accuracy of consensus* in large networked multi-agent systems. We define the level of *accuracy* as an expected probability of reaching the correct opinion by each agent given a small number of noisy observations dispersed in the system. We ensure that the number of observations is too small for a single agent to form an accurate and correct opinion. Therefore, agents have to communicate in order to improve the level of their own accuracy, and thus the accuracy of consensus, in comparison to the accuracy of introduced observations.

**Requirement 2: Minimal Communication**

Following the restrictions defined in our problem, the second requirement imposes a strict limit on communication in the multi-agent system. Specifically, we assume
that agents are only able to communicate to their peers messages comprised of their own opinions without any supporting information. We define minimal communication as the number of messages required to share a single opinion to all agents. This corresponds to a case in which the first opinion introduced into the system, which might possibly be incorrect, is shared to all agents in a single cascade. The solution for our research aim should not introduce any communication other than the opinion sharing that is already present, and should approach the level of minimal communication. As a result, individual agents must rely solely on locally available information in making their decisions. In our scenarios the only additional information that may or may not be available in a communicated message is that which might identify the sender of this message. Thus, we should consider two cases of this requirement:

**Requirement 2a: Dealing with Anonymous Peers**

This more difficult case corresponds to scenarios where the identity of a sender is concealed. For example, anonymous peer-to-peer networks, (e.g. a distributed file sharing network); or a radio communication network in which an agent broadcasts its opinion only to its nearest peers in a small surrounding area, without identifying itself in order to spare the bandwidth and its energy resources.

**Requirement 2b: Dealing with Identified Peers**

In other cases, we are often able to identify the sender of a message even when this information is not included in the message itself. For example, network connections might be fixed or each communication channel be dedicated to a specific peer. In a number of other scenarios, such as social or sensor networks, senders identify themselves before broadcasting a message to their peers. Thus, the developed methods might also benefit from identifying the sources of specific opinions and should use this assisting information.

**Requirement 3: Adaptivity and Scalability**

The agents in the described scenarios typically operate in environments with the complex topological properties of interactive networks, often with thousands or millions of peers. Such topological properties have been shown to have a significant impact on the dynamics of opinion sharing processes (Watts and Strogatz, 1998). Considering this, in order to operate in different settings, solutions to our problem have to be: (i) adaptive to system parameters, and therefore should not require specific tuning for any given domain; (ii) scalable to large system sizes by efficiently inducing the desired mode of behaviour in a limited time. Previous research also suggests that small multi-agent systems with less than hundreds of agents do not exhibit stable properties of collective behaviour (Glinton et al., 2009) and, therefore, we should also investigate the lower boundary of acceptable numbers of agents for our solution.
Chapter 1 Introduction

Requirement 4: Robustness and Flexibility

Our solutions should meet the requirements mentioned above even when a significant number of agents do not participate in a process of self-organising the system into a desired mode. This will enable the application of such solutions in large heterogeneous systems, such as human-agent or sensor networks, where it might be impossible to update behaviour of all the agents. Moreover, robustness is essential to mitigate any negative effects introduced by the agent failures that are likely to occur in large multi-agent systems.

The model that is closest to our requirements was offered by Glinton et al. (2010b). This, however, suffers from a number of shortcomings, which we have briefly discussed above. Therefore, to meet our research requirements, we need to develop and analyse an appropriate model of our problem. Despite this, analysis of opinion formation has already lead to the development of the DACOR algorithm Glinton et al. (2010b) which is a promising solution to meet our Requirement 1. However, neither this solution nor any of the existing work to date meets all the requirements together, which is necessary to make a progress in this area. Thus, we address this research gap and in the next section describe our specific contributions as presented in this thesis.

1.5 Research Contributions

To achieve our aims, we designed a new model of opinion sharing and two solutions that significantly improve the accuracy of consensus. By analysing our model, we confirm the presence of different modes of collective behaviour, which depend on the weights agents attribute to their peers. These weights encode the relative influence of individual peers’ opinions on an agent’s own belief. When weights are tuned into a narrow range, which is highly dependent on the system’s parameters, the accuracy of consensus significantly increases.

Our solutions to the research problem tune the weights that agents attribute to each other in order to self-organise the system into the desired behavioural mode. In view of our research requirements, these solutions are designed as decentralised behavioural algorithms that curate the actions of each agent individually, given only their local views. On a large scale these algorithms steer the whole system into a self-organised mode in which opinions are aggregated in a distributed fashion and thereby the accuracy of consensus significantly improves.

In more detail, we advance the state-of-the-art in the following ways:

1. **Opinion Sharing Model with the Gradual Introduction of Observations**

   In order to measure the accuracy of consensus, we present a model that simulates
the gradual introduction of conflicting opinions into a system. In comparison, the well-known opinion formation models mentioned above initially endow agents with opinions, which dramatically changes the opinion dynamics.

Our model is the first to quantify the impact of collective behaviour on the level of accuracy achieved. This model addresses the shortcomings of the existing model offered by Glinton et al. (2009) by incorporating a new process to gradually introduce new opinions into the system. This enables us to quantify the accuracy of consensus, analyse the rate of convergence to a consensus and, most importantly, derive analytical bounds on the performance metrics. Additionally, making changes to the model enables us to study its behaviour with alternative decision rules being employed by the agents. Crucially, this enables us to show that an increase in accuracy comes from exploiting the properties of collective behaviour regardless of the specific agent design, as long as they can tune the weights they attribute to each other.

2. Accurate Consensus with Anonymous Peers

We exploit the properties of the collective behaviour in the model to design the first algorithm that meets our research requirements in the case when peers are anonymous (Requirement 2a). Specifically, we develop a novel decentralised algorithm, Adaptive Autonomous Tuning (AAT), that improves the accuracy of consensus in a large multi-agent system with a complex communication network. It does so by tuning the weights of each agent individually and self-organising the system into the critical mode of collective behaviour. In this mode, the multi-agent system filters early and possibly inaccurate opinions by sharing them amongst small groups of neighbouring agents to prevent overreaction. Only when several groups with the same opinion overlap is this locally supported opinion disseminated on a large scale, thereby leading to the consensus.

Crucially, AAT is the first solution that meets the minimal communication requirement. In contrast, the existing state-of-the-art algorithm, DACOR, is a message-passing algorithm that communicates 4-7 times more service messages than is required to share decentralised opinions.

Moreover, we empirically evaluate AAT and show that it significantly outperforms DACOR. Specifically, using AAT, the accuracy of consensus reaches 82-93% given only 5% of agents with noisy sensors (which make only 65% of observations corresponding to the correct state), while the remaining 95% of the agents do not have direct access to the observations. This figure is significantly higher than the 70-75% reached by DACOR and close to the 94-97% attained by systems pre-tuned for the highest accuracy by an intensive empirical exploration of its parameters. Finally, AAT has lower computation costs and requires up to $5 \cdot 10^4$ times fewer agent actions than DACOR to achieve the beneficial self-organised mode.
Furthermore, we show that AAT is the first decentralised solution designed to improve the accuracy of consensus in heterogeneous systems which include faulty or indifferent agents that do not participate in the weights optimisation process. Specifically, it significantly improves accuracy when up to 80-90% of the agents in the system are not controlled by AAT. This implies that AAT can potentially be introduced into existing large systems where it is impossible to update the behaviour of all their agents simultaneously.

This work is discussed in Chapter 4 and has led to the following publications:


3. Accurate Consensus with Identified Peers

Following the above, we investigate how we can benefit from peer identification and design our second solution to meet our research requirements in a case when peers are known (Requirement 2b). Specifically, the Individual Weights Tuning (IWT) algorithm is the first solution which differentiates the peers of an agent and adjusts the individual weights attributed to their opinions.

In developing this algorithm, we explore when it is beneficial for an agent to differentiate its peers and then analyse how agents can identify the most influential peers given only the history of their opinion sharing. We investigate a number of metrics and come up with an adaptive solution that does not rely on external parameters. As a result of this, IWT meets the performance of AAT in sparse networks, and crucially, results in a significantly higher accuracy of consensus in dense and scale-free networks. This contribution is presented in Chapter 5.

Next, we describe the structure of this thesis by outlining the content of the following chapters.

1.6 Thesis Structure

The following list outlines the structure of the remaining chapters:

---

2http://eprints.soton.ac.uk/272435/
3http://eprints.soton.ac.uk/273087/
In Chapter 2 we provide a review of the related literature with regards to the research aims and requirements. This survey covers the areas of modelling large multi-agent systems; dynamic processes in these systems, such as the spread of diseases, norms and opinions; and interconnection between different models. Then we discuss the collective behaviour that arises from the interaction of individual agents. To explain this in detail, we analyse how system properties influence this behaviour, such as the underlying network topology that defines communications links between agents and agents’ decisions rules which curate their individual behaviour. Finally, we show how the predicted collective behaviour can be exploited in order to improve the overall accuracy of consensus in opinion sharing scenarios. We also find that the existing state-of-the-art solution, the algorithm DACOR, does not meet our research requirements and we develop new approaches in the following chapters.

In Chapter 3 we explain the shortcomings of existing models and introduce our model of sharing conflicting opinions in large multi-agent systems. Following this, we analyse its behaviour, introduce benchmarks, offer a number of possible experimental setups and perform its empirical evaluation. By doing so, we identify the properties of collective behaviour that improve the accuracy of consensus.

In Chapter 4 we present our decentralised algorithm, Autonomous Adaptive Tuning (AAT), that improves accuracy of consensus by exploiting the properties of collective behaviour. We demonstrate that AAT outperforms the existing state-of-the-art algorithm, DACOR, and that it is the first to meet our research requirements. However, AAT is designed for setting in which agents do not differentiate their opinion sources, while in many other opinion sharing scenarios agents can learn who their sources are.

In Chapter 5 we present our Individual Weights Tuning (IWT) algorithm, which addresses the described gap and benefits from additional information. Specifically, using IWT agents differentiate their peers and assign them with individual weights according to their preferences. By doing so, IWT improves the accuracy of the consensus even further than AAT and in even more challenging settings, such as dense networks.

Finally, Chapter 6 concludes with a summary of our research and the outlook for future work.
Chapter 2

Related Work

In this chapter we review the relevant literature. By bringing together advances from the fields of multi-agent systems, complexity science and some aspects from other research fields, we can narrow down our research problem to the specific task that has to be solved. In order to do so, we discuss the problem of forming a consensus in large systems and how it is approached in existing biological, social or even physical systems. This gives us understanding of the dynamic processes in such systems, which of their aspects are crucial in modelling and their expected influence on the accuracy of consensus. Then, in Section 2.2, we show how such large multi-agent systems are modelled and analysed. We discuss how macro-level patterns and overall complexity arise from local interactions between agents, a number of the diverse models that have been developed to explain such patterns and the relevance of their findings to our topic. Section 2.3 summarises the existing models of opinion sharing in large systems, assumptions that were made in their development and how closely they compare to the settings of our problem. Importantly, these models predict the influence of dynamic processes on the accuracy of opinion consensus, and in the following section we show the first decentralised algorithm to benefit from this. Finally, in Section 2.5, we summarise the chapter and point to gaps in the existing research that has to be addressed in order to solve the problem of improving the accuracy of consensus in large multi-agent systems.

2.1 Consensus in Large Systems

When a group of individuals come together, they face the problem of efficient communication in order to make the most from being in the group. For example, this is essential for individuals to establish norms and reach agreements in a society; to make group decisions on directions of movement in a flock of birds, school of fish or swarm of bees; or to reach a group agreement in the presence of faults in a distributed computing system. Quite often the essence of these problems, which we discuss later in detail in our
Chapter 2 Related Work

motivating scenarios, is to reach the *consensus* that is a general agreement in opinion between the individuals in the group (Merriam-Webster, 2012). Here, the *opinion* of each individual is its subjective statement about a common issue or subject of interest. In our research we assume that the common subject of interest has a single true state and that diversity of opinion comes from observational uncertainty. This assumption enables us to reason about the accuracy of consensus which is the expected probability of an individual forming the *correct opinion* that corresponds to the true state. In essence, our research aim is to improve the accuracy of consensus. However, before we approach the challenging settings of our problem, we briefly overview how consensus may be reached and how its accuracy is analysed.

In the diversity of scenarios we mentioned above, opinions may be formed as the result of the perspective of an individual, its observations and interpretations, and finally, its particular feelings, beliefs, and desires. Regardless of how individuals gather information and form their initial opinions, when they are part of a group they must translate everyone’s opinions into some form of consensus. There are a number of ways to make the transition from diversity of opinion to consensus. Here we briefly discuss existing methods. In doing so, we approach our research problem by gradually increasing the complexity of the settings in which groups operate.

If a group is under the control of a *centralised authority*, such as a major news agency that gathers reports from diverse sources, the process of reaching consensus is straightforward. There are two basic options for aggregating opinion diversity: (i) to form an *opinion supported by a majority* (known as the *majority rule*); or (ii) determine an *average opinion*. The latter method is appropriate when we deal with *continuous opinions*, such as individuals’ estimates of the weight of an ox, which is known as a state estimation problem. Specifically, a classic experiment showed that the average guess of nearly 800 people provides a very accurate estimate of the weight of an ox (Galton, 1907). As long as reported opinions are independent, the opinion formed by a centralised authority is expected to be much more accurate than the accuracy of a single individual. This effect, when a group as a whole outperforms most of its individuals, is known as ‘the wisdom of crowds’ (Surowiecki, 2004). Increase in accuracy and the key role played by diversity of opinion is explained by the diversity prediction theorem (Mason and Page, 2007; Page, 2008). Given a number of opinions $o_i, i \in 1 \ldots n$ about the true state $\theta$ of the subject of observations, the error of an average opinion $c$ is the following:

\[
\text{Collective error} = \text{Average individual error} - \text{Opinions diversity}
\]

\[
(c - \theta)^2 = \frac{1}{n} \sum_{i=1}^{n} (o_i - \theta)^2 - \frac{1}{n} \sum_{i=1}^{n} (o_i - c)^2
\]

(2.1)

It shows that the collective error of the group is smaller than the average individual error because of the diversity of opinion. Most importantly, this result confirms that
individuals generally benefit from joining a group by forming a more accurate opinion than they would have formed on their own.

However, it is clear that the result of this aggregation depends on the specific averaging technique, which might be a mean or median of the reported opinions, or even a weighted aggregation given confidence in specific reporters. These different domain-specific averaging techniques result in distinct group dynamics which we will need to focus later on. Therefore, in our research we analyse groups with discrete opinions which represent a number of possible alternatives regarding a subject of interest of which only one is correct. For example, a weather forecast reports only one possible state from a set of alternatives: \{sunny, cloudy, rain, fog, ...\}, similarly, following an earthquake, a sensor network designed to predict a tsunami must make a discrete \{yes, no\} decision on issuing a warning. To appropriately aggregate discrete opinions we need to apply a threshold rule, such as the majority rule we discussed above. Since we are not able to measure the distance between different discrete opinions and the correct opinion, we cannot measure the result with precision, and thus, the collective error cannot be defined. Due to this, we offer the accuracy metric as the probability of forming the correct opinion, which we will use from now on.

2.1.1 Accuracy of Consensus

A remarkable increase in accuracy of consensus compared to an individual’s accuracy was documented in prediction problems solved by a group of people (Mason and Page, 2007). This phenomenon, named ‘group intelligence’ (Fisher, 2009), was explained in the 18th century when a democratic process of voting was mathematically justified by Condorcet’s jury theorem (Boland, 1989). The theorem states that if each member of a group has a better chance than 0.5 out of \([0 \ldots 1]\) of forming the correct opinion, then the accuracy of a majority consensus rapidly becomes closer to ideal 1 as the size of the group increases. If we denote the accuracy as $R_{\text{CJT}}$ and the size of a group as $N$ in which all individuals have the same probability of forming the correct opinion $r_i = r \ \forall i \in 1 \ldots N$, then Condorcet’s jury theorem can be defined based on a cumulative function of the binomial distribution:

$$R_{\text{CJT}} = \sum_{k=\left\lceil \frac{N}{2} \right\rceil}^{N} \binom{N}{k} r^k (1 - r)^{N-k}$$

where factor $\left\lceil \frac{N}{2} \right\rceil$ defines the number of opinions required for a majority. Even if most if individuals form a random opinion ($r_i = 0.5$), a few knowledgeable experts ($r_i > 0.5$) are able to lead the whole group into the correct consensus. In the following Chapter 3 we use this result as a benchmark to show how a group can perform in ideal settings. Specifically, this theorem assumes that individuals are independent, which means that they do not influence each other’s opinions. However, in realistic settings this influence...
is present in a forms of local communication between neighbouring individuals, resulting
in a network of social ties and leading to within-group correlations.

Crucially, the majority rule assumes that there is a central authority that simultaneously
aggregates the opinions of all the individuals in order to make a decision on consensus.
Therefore, it is not applicable in the distributed and networked groups which we de-
scribed in our motivating scenarios. A large body of other voting techniques has been
studied besides the majority rule (Shoham and Leyton-Brown, 2008, Chapter 9), but
they also require a central authority in order to operate. Thus, we need to look for other
methods to reach consensus in decentralised systems.

2.1.2 Consensus in Decentralised Systems

There are many natural systems with a large number of individuals which have to reach
consensus in order to prosper. Well-studied examples are honeybee swarms choosing a
new nest (Visscher et al., 2006) and flying towards it (Beekman et al., 2006), marching
locusts (Buhl et al., 2006), and movements of flocks of birds or schools of fish (Simons,
2004; Ward et al., 2008; Sumpter et al., 2008). Most of them share the same properties
as our motivating scenarios. Specifically, in contrast to the voting techniques, these nat-
ural systems are decentralised, networked groups with limited communication between
their individuals. Individuals in such systems are able to interact only with their nearest
peers or network neighbours. To achieve consensus in such settings and to improve their
chances of forming the correct opinion, animals copy opinions from their neighbours
(usually in the form of a direction of movement or a choice from a number of alterna-
tives). Complexity science, which we briefly discussed in the introduction, has shown
that collective behaviour in animal groups emerges from a set of very simple rules of in-
teraction between neighbours. The imitation rule of copying neighbours’ opinions is one
example. A rise in consensus is one of the outcomes of collective behaviour introduced
by this rule. In contrast to the ‘group intelligence’ concept with a centre responsible
for consensus formation, this emergent property of forming consensus given only local
interactions is called ‘swarm intelligence’ (Fisher, 2009). It allows a group to tackle and
solve problems in a way that its individual members cannot. Specifically, a group is able
to respond collectively by aggregating opinions from a number of informed individuals,
such as scout honeybees in a swarm looking for a new nest site or fish in the front of a
school, and sharing them to the rest of the group in a decentralised fashion.

When individuals can only interact with their neighbours opinion sharing occurs as a
wave of rapid propagation from individual to individual, which is one of the key features
of collective or ‘swarm’ behaviour. This chain reaction is known as cascading behaviour
and we focus on its analysis later. The cascade effect in groups of animals is known as
quorum response (Sumpter and Pratt, 2009). The group arrives at a consensus in which
each individual’s likelihood of choosing an opinion increases steeply with the number of
neighbours already committed to that opinion. Voting techniques require opinions to be independent in order to form the most accurate consensus, while in quorum response individuals in a group exhibit interdependence of their opinions due to the nature of peer to peer sharing. The same opinion may arrive through different paths or even re-affect the source and make a group overconfident. This points to a significant problem with quorum responses, in that individuals must choose whose opinions to copy.

Our general understanding of such complex patterns of behaviour come from: (i) observations of the real world (groups of animals, humans and some physical systems); (ii) designing abstract formal models to conceal and independently analyse key properties and; (iii) developing agent-based models for computer simulations. The latter approach, which we adopt in our work, combines accurate description of complex settings with tools for conducting controlled and repeated experiments. In the following sections we discuss the design of a multi-agent system for our research problem.

2.2 Modelling Large Multi-Agent Systems

The basic constituents of the social phenomena of opinion sharing, that we discuss in this thesis, are agents that interact with a limited number of peers. This number of peers is usually negligible when compared to the total number of agents. Due to a large number of participants, such networked societies are characterized by stunning global regularities (Ball, 2005). They exhibit transitions from disorder to order, like the emergence of consensus about a specific issue when a majority of the agents share the same opinion. Early results in this field show that the dynamics of such systems cannot be explained just in terms of their simplistic elements (Anderson, 1972). As we discussed in the Introduction chapter, such emerging macroscopic phenomena were initially analysed from the perspective of statistical physics in an attempt to understand regularities at a large scale, such as the collective effects of interaction among single individuals (Castellano et al., 2009). However, this approach implies that individual agents are simple entities and that the models used to describe social systems are too simplified to describe any real situation, such as the highly-acclaimed models by social scientists of urban segregation (Schelling, 1971), cultural dissemination (Axelrod, 1997) and the Voter model of opinion formation (Krapivsky, 1992; Frachebour and Krapivsky, 1996). Nowadays, with access to modern computational resources, a vast array of agent-based models can be simulated in more complex settings and large systems can be analysed in much finer detail. Together, these two approaches from the perspective of statistical physics and agent-based simulations enable us to combine recent theoretical findings that predict the emergence of desired properties, and develop and evaluate an agent-based solution for our research problem.
Chapter 2 Related Work

However, the properties of the communication network have a significant influence on sharing processes (Boccaletti et al., 2006). As the result, the properties of the network influence the range of parameters when different modes of collective behaviour are observed. Therefore, our first step in decomposing the research problem is to analyse the environment that is the structure of a multi-agent system.

By discussing our motivating scenario in Chapter 1, we showed that relations between agents form complex communication networks. For example, Figure 2.1 depicts a part of a real business-oriented social network which only shows relations in the neighbourhood of one individual. Here, nodes represent people and links represent business ties between the people. Highly interconnected people from different societies form a number of groups with which this individual is involved (marked with different colours). Clearly, due to the dense connections in the separate groups, members of such groups are more likely to share common beliefs and opinions, and the most influential members are the individuals with the largest number of ties (marked by the size of a node).

![Figure 2.1: Example of a business-oriented social network which only shows relations in the neighbourhood of one individual. The map is colour-coded to represent different affiliations or groups from the individual’s professional career, such as previous employers, college classmates or industries the individual has worked in. Bigger nodes represent people who are the most connected within that specific cluster or group.](http://inmaps.linkedinlabs.com/)

As we can see, even a small part of a much larger social network can exhibit complex relations between individuals and therefore we need techniques to describe these properties. Moreover, it was shown that the properties of communication networks are fundamentally linked in models of social dynamics, including opinion sharing models (Keeling and Eames, 2005).

More specifically, in this section we introduce definitions and notations, and discuss the basic quantities used to describe the topology of the network. Then, we analyse properties observed in real networks, and provide a brief review of the modelled topologies motivated by empirical observations.
2.2.1 Structural Properties of Decentralised Systems

We rely on network analysis that can provide a wealth of quantitative and qualitative information about social network connections. Specifically, network analysis has been used as an explanatory tool to describe the evolution and spread of ideas and innovations in societies (Leinhardt, 1977), and observed social dynamics can often be understood through analysis of the social networks that underlie them.

Historically, communication networks have been studied by a branch of discrete mathematics known as graph theory. Graph theory is the natural framework for the exact mathematical treatment of complex networks and, formally, a complex network can be represented as a graph (Mesbahi and Egerstedt, 2010). Graph-based abstractions of networked systems contain no information about what exactly is shared between agents, through what protocol the exchange takes place, or what is subsequently done with the received information. Instead, the graph-based abstraction contains high-level descriptions of the network topology in terms of nodes and edges that represent the agents and their communication links respectively.

In order to describe the relations between the agents, we introduce notation based on graph theory which will be used throughout this thesis. By a large multi-agent system we mean a large set of agents that are connected by a number of communication links. We define agents as:

\[ A = \{ i_1, \ldots, i_N \}, \quad N = |A| \gg 100 \]  

(2.3)

where \( i \) is an agent or a node in the graph. The size of a system \( N \) has to be large to meet our research requirements and also to decrease the relative contribution of a single agent into the system’s dynamics. Later, this enables us to focus on the influence of dynamic processes on the accuracy of consensus. This allows us to use a simplified model of an agent without losing important system-wide properties. Finally, as we discuss later, in some settings a large multi-agent system can exhibit macroscopic phenomena of social behaviour that are not present in individual relations and thus, do not stand out in small systems.

The communication links, or the edges between agents are denoted as a set of possible links:

\[ E = \{ (i, j) : i, j \in A \} \]  

(2.4)

where each connection \((i, j)\) corresponds to the link between agents \( i \) and \( j \), which are said to be connected and referred to as neighbouring. Connections are often reciprocal and undirected, and agents can pass pieces of information either way across a link. These settings can be found in most social networks of rumour spreading, collaboration networks and in structures of online societies such as Facebook. However, this is not necessarily always the case and in some systems, such as news dissemination networks, an agent can have a large number of subscribers and information can only travel one way.
along a link. Our own empirical evaluation showed that the key properties of the model discussed later in this chapter, do not change qualitatively when links are directed or undirected. Therefore, to simplify the notation and analysis, we make an assumption that links are undirected, i.e. \((i, j) = (j, i)\). Additionally, we assume that connections do not introduce restrictions on information that is shared over them, and in terms of graph theory they are unweighed. These assumption are in place in all the models we discuss in the following sections.

Together the two sets, \(A\) and \(E\), form an undirected graph that represents the structure of the large system \(G(A, E)\). Due to the focus of our research on sharing processes in a system, we assume that the graph is connected and that an agent’s opinion can reach any other agent by following the network links. Specifically, this implies that for any two agents \(i\) and \(j\) in the system there exists a number of intermediate neighbours \(l_{i,j}\) that can pass an opinion between them. This corresponds to the concept of path, that is the natural distance measure between two nodes, and is defined as the number of nodes traversed by the shortest connecting path. This distance \(l_{i,j}\) is called the shortest path length and in the connected network it is finite \(0 \leq l_{i,j} < N\), and it is symmetrical for undirected graphs \(l_{i,j} = l_{j,i}\). The effective definition of the linear size of the network is the average shortest path length, defined as the average value of \(l_{i,j}\) over all the possible pairs of nodes in the network:

\[
\langle l \rangle = \frac{1}{N(N - 1)} \sum_{i,j \in A, i \neq j} l_{i,j} \tag{2.5}
\]

This measure can be used to compare networks with different topologies in terms of how fast a single opinion can be shared within them.

Another important feature of graphs which helps in understanding generic properties of their structure, is their sparseness. The number of edges \(|E|\) for a connected graph ranges from \(N - 1\) for nodes connected in a line, to \(\left(\frac{N}{2}\right)\) for a fully connected network. There are different definitions of sparseness, but we adopt the convention that when the number of edges scales as \(|E| \sim N^\alpha\) with \(\alpha < 2\), the graph is said to be sparse (Barrat et al., 2008). In the case where \(|E| \sim N^2\), the corresponding graph is called dense. However, as we identified in the discussion of our research requirements, agents are often limited in their computation and communication capabilities. Therefore, agents in large systems often form sparse networks, in which they have to pass opinions through intermediate peers to inform the whole system. This passing of opinions dramatically changes the dynamic processes, and as we we see later in Chapter 5, leads to different solutions.

When looking at networks, one of the main insights is the importance of their basic elements. The importance of an agent is commonly defined as its centrality. We focus on degree centrality or, simply, the degree of an agent. The degree \(d_i\) of agent \(i\) is defined as the number of edges in the graph incident on the node \(i\) or, in other words, as the
number of its neighbours. The set of neighbours, or the neighbourhood, is defined as the agents that have links to agent $i$:

$$D_i = \{ j : \exists (i, j) \in E \}$$  \hspace{1cm} (2.6)$$

Considering this, we can define the degree as $d_i = |D_i|$. The degree of a node has an intermediate interpretation in terms of the centrality measure quantifying how well an agent is connected to other agents in the graph. However, in large systems, regularities cannot be detected by looking at local elements and their properties. In other words, we have to shift our attention to statistical measures that take into account the global behaviour of these quantities. Specifically, we can define the degree distribution as $P(d)$ that is the probability that any randomly chosen node has degree $d$. It is obtained by constructing the normalised histogram of the degree of the nodes in the network. Consequently, the average degree is the average value of $d_i$ over all the nodes $i$ in the network and can be defined as:

$$\langle d \rangle = \frac{1}{N} \sum_{i \in A} d_i = \sum_d d P(d) = \frac{2|E|}{N}$$  \hspace{1cm} (2.7)$$

A sparse graph has an average degree that is much smaller than the size of the graph, i.e. $\langle d \rangle \ll N$. In the following sections we show that the properties of the degree distribution, $P(d)$, are crucial in identifying different classes of networks.

Along with the degree measures, nodes are characterized by the structure of their local neighbourhood. For example, in a spatial network, such as a road network, it is quite possible that two neighbours of an agent are connected to each other. This property is referred to as the clustering coefficient, $C(i)$, which is defined as the average fraction of pairs of neighbours of node $i$ that are also neighbours of each other, and which measures the local group cohesiveness (Watts and Strogatz, 1998). Given agent $i$, the clustering coefficient $C(i)$ is defined as the ratio of the number of links between the neighbours of $i$ and the maximum number of such links:

$$C(i) = \frac{|\{(j,l) \in E : j,l \in D_i\}|/2}{d_i(d_i-1)/2}$$  \hspace{1cm} (2.8)$$

where the numerator measures how many of agent $i$’s neighbours have connections between them, and the denominator represents the maximum number of connections if all neighbours are linked. In Figure 2.2, we provide an illustration of some simple examples of the clustering of vertices within a given neighbourhood. The average clustering coefficient of a graph is simply given by:

$$\langle C \rangle = \frac{1}{N} \sum_{i \in A} C(i)$$  \hspace{1cm} (2.9)$$

Apart from the most essential concepts we introduce here, there is a large number of
other topological properties that we do not consider in this thesis. To omit such deep analysis of network systems, we rely on existing studies and in the following sections draw out the most important conclusions about existing topologies that are relevant to modelling our problem.

2.2.2 Networks of Interactions

Although the theoretical representation of networks was initially introduced by Euler in the 18th century, our understanding of complex structures in real networked systems has formed only during the last two decades. This recent change was facilitated by growth in the availability of large database and computing facilities, as well as the development of powerful and reliable data analysis tools. All these advances have constituted better machinery for exploring the topological properties of networked systems from the real world. This has allowed the study of the topology of interactions in a large variety of systems, and specifically, in our field of interest, communication (Faloutsos et al., 1999; Pastor-Satorras and Vespignani, 2004) and social networks (Watts and Strogatz, 1998; Leskovec et al., 2009; Barrat et al., 2008, chap. 2). The main findings of research in this area are that despite the inherent differences, most real networks are characterised by the same topological properties, such as relatively small average path length $\langle l \rangle$, a high clustering coefficient $\langle C \rangle$, a fat-tailed shape in the degree distributions $P(d)$ and emergent community structures (Boccaletti et al., 2006). All these features make real networks radically different from the standard models studied in mathematical graph theory such as regular lattices and random graphs. This, in turn, has led to significant attention being directed towards understanding the evolutionary mechanisms that have shaped the topology of networks, and to the design of new models reflecting the most significant properties that are empirically observed.

In this subsection we review the basic topological features that characterise real-world networks into broad classes according to their observed statistical properties. In particular, the emergence of small-world and scale-free properties are discussed as prominent concepts which have led to a paradigm shift in which the classification and modelling
of the processes of network formation have become a central issue. As a result, a discussion of these findings provides us with insights into how typical topologies influence opinion sharing in real-world complex networks and suggests the most suitable models of networks to evaluate our solution with later.

2.2.2.1 Small-World Property

The earliest study of social relations by Milgram (1967) sought to determine whether most pairs of people in a society were linked by short chains of acquaintances. For this purpose a number of individuals were recruited to attempt to forward a letter to a given addressee through people they knew only on a first-name basis. Of the completed chains, the median number of required steps was six. This became known as the small-world phenomenon and entered popular culture as the principle of ‘six degrees of separation’ (Watts, 2003). Further study of dynamic processes across information systems and relations in social networks has pointed out that apart from the expected local connections in an individual’s vicinity there is a number of short-cuts. Specifically, short-cuts are bridging links that connect different areas of the networks, thus speeding up the communication among otherwise distant nodes. Thus, the connectivity in such systems exhibits the same small-world network characteristics (Watts and Strogatz, 1998).

One reason for the current empirical consensus that social networks generally are “small worlds” is that this notion has been increasingly confirmed in settings where we do have full data of the network structure. For example, experiments conducted by Dodds et al. (2003) on e-mail exchanges successfully reproduced Milgram’s experiment.

In most real networks, despite their often large size, there is a relatively short path between any two nodes. For example, Internet packages travel at most through a few dozens of routers instead of $10^3$ required for a regular grid (Faloutsos et al., 1999). This feature is known as the small-world property and is mathematically characterized by an average path length $\langle l \rangle$, defined as in Equation 2.5, that depends at most logarithmically on the network size $\langle l \rangle \sim \log N$ (Watts and Strogatz, 1998; Watts, 2004). The small-world property in real networks is often associated with high values of the clustering coefficient, defined as in Equation 2.9. For this reason, Watts and Strogatz (1998), in their pioneering paper, have proposed defining small-world networks as those networks having both a small value of $\langle l \rangle$, like random graphs, and a high clustering coefficient $\langle C \rangle$, like regular lattices. In other words, this definition indicates that such networks are extremely efficient in exchanging information both at a global and at a local scale (Latora and Marchiori, 2001).
2.2.2.2 Scale-Free Degree Distribution

Another well-studied property that affects information exchange is the degree distribution \( P(d) \), which is the probability that a randomly selected node \( i \) has exactly \( d_i \) neighbours. The usual case in mathematical graph models is to assume that a network is homogeneous, and thus, all nodes are topologically equivalent like in regular lattices or in random graphs. A regular lattice has a simple degree distribution with a single sharp spike (delta distribution). Any randomness in the network connections broaden the shape of this peak, and the degree distribution becomes a binomial or Poisson distribution in the limit of large network size. Therefore, it was expected to find in real networks the degree distribution localized around an average value. In contrast, it was found that most of the real networks display power-law shaped degree distribution

\[
P(d) \sim d^{-\gamma}
\]

with exponents varying in the range \( 2 < \gamma < 3 \) (see Figure 2.3) (Boccaletti et al., 2006). The average degree \( \langle d \rangle \) in such networks is therefore well defined and bounded. On the other hand, a measure of the typical error we make if we assume that \( \langle d \rangle \) is the typical degree value of a node approaches infinity in the asymptotic limit of infinite network sizes, so fluctuations are unbounded and depend on the system size (Barrat et al., 2008, chap. 2, app. 1).

![Figure 2.3: Comparison of Poisson (squares) and power-law degree distribution (circles). The two distributions have the same average degree \( \langle d \rangle = 10 \). The dashed line corresponds to the power law \( d^{-\gamma} \), where \( \gamma = 2.3 \) (Barrat et al., 2008).](image)

Such networks have been named scale-free networks (Barabási and Albert, 1999), because power-laws exhibit the property of having the same structural form at all scales. These networks, having a highly inhomogeneous degree distribution, result in the simultaneous presence of a few nodes (the hubs) linked to many other nodes, and a large number of poorly connected elements (see Figure 2.1 as an example of scale-free topology).

In the context of opinion sharing in a multi-agent system, the most influential opinions are introduced by the hubs in the network. This intuition was proven by the early studies of citation networks between scientific papers. In particular, Price (1965)
showed that the number of citations of papers has a distribution following a power law. The mechanism that leads to such a distribution was called “cumulative advantage” or “preferential attachment” (Price, 1976). The emergence of scale-free networks is noticed in many other areas and is also called the “rich-get-richer” or popularity phenomenon (Easley and Kleinberg, 2010, chap. 18). For example, the Web, the Internet and collaboration networks exhibit these properties where the fraction of nodes with very high degrees $d_i \gg \langle d \rangle$ is much larger than one would expect based on models of random graphs (Faloutsos et al., 1999). Therefore, there is a need to model topologies of such networks in order to study their dynamics processes.

### 2.2.2.3 Network Topology Generators

The observed statistical properties of real-world networks motivate us to choose appropriate network topology generators. In this section, we present topology generators that are used to study dynamic processes and discuss their significant properties. Specifically we discuss the following classic models and their generators: a random network, a small-world network and a scale-free network. These different models help us to determine the influence that specific network features have on the social dynamics of opinion sharing that we analyse in the following sections.

#### 2.2.2.3.1 Random Networks

The static random network and the corresponding topology generator proposed by Erdős and Rényi (1959) is the simplest network model that includes stochasticity as an essential element. It is characterised by an absolute lack of knowledge of the principles that guide the creation of connections between nodes. Lacking any information, the simplest assumption one can make is to connect pairs of nodes at random with a given connection probability $p$. In its original formulation, an Erdős and Rényi (ER) graph is constructed starting from a set $A$ of nodes which are joined by $E$ edges whose ends are selected at random among the $N$ nodes, prohibiting multiple connections. ER random graphs are the best studied among graph models. As we show later, unlike in graphs with complex topologies, processes of information dissemination in random graphs can be analytically analysed assuming homogeneity of the structure for infinitely large network size. Although they do not reproduce most of the properties of real networks discussed in Section 2.2.2, ER models exhibit an average path length that scales logarithmically with the graph size. This scaling behaviour is the signature of the small-world effect observed in many complex networks (Bollobás, 1981).
2.2.2.3.2 Small-World Networks

Although the random graph model exhibits scaling of the average path length, its clustering coefficient is determined by the imposed degree distribution and vanishes in the limit of very large sparse graphs. In contrast, empirical observation finds a large clustering coefficient in many real-world networks, and therefore there is a need to define a model in which it is possible to tune \( \langle C \rangle \) (Equation 2.9) to any desired value. Inspired by the fact that many social networks (Milgram, 1967; Wasserman and Faust, 1994) are highly clustered while at the same time exhibit a small average distance between vertices, Watts and Strogatz (1998) have proposed a model that interpolates between ordered lattices (which have a large clustering coefficient) and purely random networks (which possess a small average path length).

The Watts and Strogatz (WS) model is based on a rewiring procedure of the connections implemented with a probability \( p_{\text{rewire}} \). The starting point is a regular network with a ring topology, in which each node is symmetrically connected to its 2m nearest neighbours for a total of \(|E| = mN\) edges. Then, for every node, each link connected to a clockwise neighbour is rewired to a randomly chosen node with a probability \( p_{\text{rewire}} \), and preserved with a probability \( 1 - p_{\text{rewire}} \). Notice that for \( p_{\text{rewire}} = 0 \) we have a regular lattice, while for \( p_{\text{rewire}} = 1 \) the model produces a random graph (see Figure 2.4). For intermediate values of \( p_{\text{rewire}} \) the procedure generates graphs with the small-world property and a non-trivial clustering coefficient. Alternative generators for constructing small-world networks, based on adding edges instead of rewiring, have also been proposed.

![Figure 2.4: Small-world topology generator: transition from an ordered network to a random network via the small-world topology, where \( p \) is a rewiring probability of the edges in the initial regular network (Barrat et al., 2008).](image)

The WS model was used to study network properties as a function of the rewiring probability \( p_{\text{rewire}} \) and the network size \( N \) (Barrat and Weigt, 2000; Newman, 1999; Barthélemy and Amaral, 1999). As observed in (Watts and Strogatz, 1998), the small-world property results from the immediate drop in \( \langle l \rangle \) as soon as \( p_{\text{rewire}} \) is slightly larger than zero. This is because the rewiring of links creates long-range edges (short-cuts) that connect otherwise distant nodes. The effect of the rewiring procedure is highly nonlinear on \( \langle l \rangle \) and not only affects the nearest neighbours structure, but also opens new shortest
paths to the next-nearest neighbours and so on. Conversely, an edge redirected from a clustered neighbourhood to another node has, at most, a linear effect on \( \langle C \rangle \). That is, the transition from a linear to a logarithmic behaviour in \( \langle l \rangle \) is faster than the one associated with the clustering coefficient \( \langle C \rangle \). This leads to the appearance of a region of small (but non-zero) values of \( p_{\text{rewire}} \), where one has both small path lengths and high clustering.

The change in \( \langle l \rangle (p_{\text{rewire}}) \) was analysed from different perspectives (Barrat and Weigt, 2000; Newman, 1999; Barthélemy and Amaral, 1999). Specifically, studies of the diffusion of knowledge in such networks show that the steady-state level of average knowledge (mean level over all agents) is maximal when the structure of the network has small-world properties, specifically, when most connections are local with roughly 10 percent of them being long distance: \( p_{\text{rewire}} = 0.09 \ldots 0.14 \) (Cowan and Jonard, 2004). We rely on this finding to generate WS networks for our empirical evaluation.

### 2.2.2.3.3 Scale-Free Networks

The large amount of work on the characterization of the topological properties of real networks has motivated the need to construct graphs with power law degree distributions. Here we discuss a class of models which reproduces the topological properties of systems as we see them today by modelling the growth processes taking place in real networks. We concentrate primarily on the model of network growth proposed by Barabási and Albert (1999), and on its variants.

The Barabási-Albert (BA) model is a model of network growth inspired by the formation of the Web and is based on two basic ingredients: growth and preferential attachment. The basic idea is that in the Web, sites with high degrees acquire new links at higher rates than low-degree nodes. More precisely, an undirected BA graph is constructed from \( m_0 \) isolated nodes, at each time step \( t = 1, \ldots, N - m_0 \) a new node \( j \) with \( m \leq m_0 \) links is added to the network. The probability that a link will connect \( j \) to an existing node \( i \) is linearly proportional to the actual degree of \( i \):

\[
P((j, i)) = \frac{d_i}{\sum_{l \in A} d_l}
\]

Because every new node has \( m \) links, the network at time \( t \) will have \( N = m_0 + t \) nodes and \( |E| = mt \) links, corresponding to an average degree \( \langle d \rangle = 2m \). The BA model is similar to a model developed by Price (1976) to explain the power law of the topology of citation networks (Price, 1965) and the power laws that appear in the distributions of cities by population (Simon, 1955).

However, the BA model lacks clustering properties and this was addressed in more recent research. The simplest solution is to embed a triangle-generating protocol into
the BA model Holme and Kim (2002). In this Holme-Kim (HK) model, new nodes are added, each with $m$ links, as in the BA model, and connected either to a neighbour of a previously connected node or by using the usual preferential attachment rule as in the BA model. The HK model produces scale-free degree distributions and a clustering coefficient that can be varied up to 0.5, and is considered to reflect realistic networks better.

To summarise, in this section we have discussed the structural properties of large decentralised multi-agent systems. Specifically, we reviewed a number of the properties of underlying communication networks and identified their influence on dynamic processes. Since the network properties are highly interdependent, we selected three widely recognised network models and their corresponding generators to evaluate the adaptivity of our solutions later. Now, we discuss dynamic processes in networked societies and specifically, opinion sharing in large multi-agent systems.

### 2.2.3 Dynamic Processes on Networks

The ultimate goal of the study of the structure of networks is to explain the behaviour of systems built upon those networks. Thus, the next step after reviewing the models of network structures is to look at the models of social processes going on those networks. Progress on this front has been slower than progress on understanding network structure, due to an imbalance in the early work between empirical evidence and theoretical modelisation. However, some important advances have been made, particularly in the study of epidemic processes in networks, rumour-spreading, and information and opinion sharing, that are relevant to our research.

When agents are connected in a network it becomes possible for them to influence each other’s beliefs and, as a result, their behaviour and decisions. In this section, we explore how this basic principle gives rise to a range of social processes in which networks serve to aggregate individual behaviour and thus produce population-wide, collective outcomes. There is a nearly limitless set of situations in which agents are influenced by others and in this thesis, we specifically focus on the influence of the opinions they hold about the common subject of interest. We discuss principles of these sharing processes and identify properties that will be used in developing our approach. Specifically, in the following section we provide a short introduction to the theory and modelling of social dynamics in networks.

In more detail, a common theme in social dynamics is the understanding of the transition from an initial disordered state to a configuration that displays order (at least partially). Such transitions have been studied abundantly in statistical physics, and methods developed in this field were employed for analysing models of social dynamics (Castellano et al., 2009). However, such mean field approach (Flyvbjerg et al., 1993)
requirements us to simplify the description of a system by approximating the behaviour of a large number of agents with a single averaged effect. This requires us to introduce assumptions that are unrealistic for our research scenarios, such as homogeneity of agents in the system and its structure (such as a regular lattice in an infinitely large random network). In contrast to this traditional approach of statistical physics, the recent development of computer simulations now plays an important role in the study of social dynamics for complex systems that cannot be averaged. Specifically, this enables the analysis of a new class of systems with complex network topologies and provides results for the existing models in much finer detail. One of the most successful methodologies used in social dynamics is agent-based modelling (Mesbahi and Egerstedt, 2010). The idea is to construct the computational devices (known as agents with some properties) and then simulate them in parallel to model the real phenomena.

The history of agent-based models can be traced back to cellular automata, that comprise systems in which each node of the network represents an agent that can be in only one of a finite number of states. These models assume that time is discrete and that, at each time step, the next state of each agent is computed as a function of its state and of the states of its neighbours on the network. The formalism for cellular automata was introduced by Von Neumann and Burks (1966) as a framework to study the process of reproduction and it is considered as the simplest representation of a complex system (Wolfram, 1994). More recently, this approach was developed and forms the research field of multi-agent systems that is concerned with systems composed of multiple-interacting autonomous agents with local views in a decentralised environment (Wooldridge, 2002). In our work we use an agent-based model of social dynamics, so it is worth summarizing some of the important concepts and tools used in this context.

In the main modelling scheme that was adopted to deal with dynamic processes in networks, we identify each node of the network with a single individual or element of the system, or agent. A dynamic description of the system can be achieved by associating each agent $i$ with a corresponding variable $o_i$ characterizing its dynamic state. The variable $o_i$ may describe a particular attribute of the agent. A typical example is in the spread of an epidemic where the variable $o_i$ indicates if the individual is healthy or infected by a given disease. In the following section we discuss a case in which $o_i$ defines an agent’s opinion. Without losing any generality, we can enumerate all possible states $o_i = \{1, 2, \ldots, l\}$ for each agent, and the knowledge of the variable state of all agents in the network therefore defines the microscopic state of the whole system. In other words, we can denote a particular configuration of the network at time $k$ by the vector $o^k = \langle o^k_i, i \in A \rangle$, where the index $i$ runs across all the agents of the network of size $N$.

To illustrate the aims of the investigation of social dynamics, we use a paradigmatic example of order-disorder transitions in physics, the one exhibited by the Ising model for ferromagnets (Binney et al., 1992). The motivation behind studies of the Ising model on networks is usually either that they can be regarded as simple models of opinion
formation in social networks (Young, 2006) or that they provide general insight into the
effects of network topology on phase transition processes. The Ising model consists of a
set $A$ of spins (agents) and each agent $i \in A$ holds a spin $o_i$ that can assume two values
$o_i = \{1, -1\}$. Each spin is energetically pushed to be aligned with its nearest neighbours.
Ferromagnetic interactions in a number of simulation steps $k$ drive the system toward
one of the two possible ordered states, with all positive or all negative spins, in the
state of the system, $o^k = o_1^k = o_2^k = \ldots = o_N^k$. At the same time, thermal noise
injects fluctuations that tend to destroy order. For low temperature $T$, the ordering
tendency wins and long-range order is established in the system, while above a critical
temperature $T_c$ the system remains macroscopically disordered. This kind of transition
is exhibited by a variety of systems and finding them is the key aim in studying models
of social dynamics.

Apart from the Ising model, which might be seen as simplistic, well-studied models of
social dynamics are the epidemic and rumour-spreading models. These two classes of
processes are radically different:

- **Epidemic spreading** has to do with the modelling of the spread of a particular
infectious disease in a population, with the aim of reproducing the actual dynamics
of the disease, and designing strategies to control and possibly eradicate infection.

- In **rumour spreading**, instead, one wants to spread the “rumours” as fast and
efficiently as possible, not to prevent them from spreading. Practical examples
are the design of protocols for data dissemination on the Internet, or strategies of
marketing campaigns. In such cases, and in contrast to epidemic spreading, one is
free to design the rules of the dynamics in order to reach the desired result.

However, there are clear connections between epidemic disease and the spreading of
rumours through social networks. Both diseases and rumours can spread from person
to person, across similar kinds of networks that connect people and, in this respect,
they exhibit very similar structural mechanisms — to the extent that the spread of
ideas is often referred to as “social contagion” (Burt, 1987). The biggest difference
between biological and social contagion lies in the process by which one person “infects”
another. With social contagion, people make decisions to adopt a new rumour (idea
or innovation), and the model of opinion formation described later focuses on relating
such underlying decision-making processes to larger effects at the network level. With
diseases, on the other hand, the process of sharing is sufficiently complex, and thus, in
the early models it was assumed to be random.

Despite the apparent diversity between the models of social dynamics, they are actually
closely connected given the methodologies they employ and, more importantly, the
general phenomena observed (Castellano et al., 2009). Opinions, cultural and linguistic
traits, social status and other phenomena are always modelled in terms of a small set
of variables whose dynamics is determined by social interaction. The interpretation of such variables is different in various cases, but often the results obtained in one case can immediately be translated to others. In all cases, the dynamics tend to reduce the variability of the initial state, which may lead to an ordered state, with all the agents sharing the same features (opinion, cultural or linguistic traits, etc.), or to a fragmented (disordered) state. Generally speaking, the drive towards order is provided by the tendency of interacting agents to become more alike. This effect is often termed “social influence” (Festinger, 1950) and can be seen as a counterpart of ferromagnetic interaction in magnets.

2.3 Models of Opinion Sharing

Social influence is at the core of social psychology and deals with the effect of other people on an individual’s thoughts and behaviours. It describes innovation adoption, decision-making, rumour-spreading and opinion formation which all unfold at a macro-level. The overarching question in these phenomena is how the micro-processes between individuals are related to the macro-level behaviour of groups or whole societies.

In particular, an important issue is understanding the diversity or uniformity of beliefs in a large number of interacting agents. If the recipient of influence usually changes its belief towards the influencer’s belief, we observe the outcome of a complete uniformity of beliefs in the system. However, this is not what we observe in reality, as minority opinions persist and we often see polarization of opinions in politics and culture. The collapse of uniformity, however, may be avoided by considering several of the other features of real-world social systems. Firstly, social influence is not always a linear mechanism. Also, the patterns of connectivity among individuals may be very complex, and foster or hinder the emergence of collective behaviour and uniformity (Barrat et al., 2008).

There are a number of opinion formation models that were developed to explain processes of social influence. The pioneering works use opinion formation models to explore how macro-level collective behaviour emerges as a function of the micro-level processes of social influence acting among the agents of the system (Granovetter, 1978; Nowak et al., 1990; Axelrod, 1997). These models adopt the statistical physics approach to explore the moments of phase transitions. Nowadays, with access to powerful computational resources, a vast array of agent-based models aimed at studying social influences have been defined and simulated to understand social behaviour in much finer detail (Castellano et al., 2009; Mesbahi and Egerstedt, 2010). A first class of models is represented by behavioural models where the attributes of agents are binary variables similar to Ising spins as in the case of the Voter model (Krapivsky, 1992), the Majority rule model...
(Galam, 2002) and the Sznajd model (Sznajd-Weron, 2000, 2005). In other cases additional realism has been introduced, such as complex topologies into the aforementioned models and models with continuous opinion variables (Deffuant et al., 2000; Hegselmann and Krause, 2002). The other development in this class of models was proposed by Axelrod (1997), in which opinions or cultures are represented by vectors of cultural traits. These models have introduced the notion of bounded confidence: an agent interacts with any other agent only if their opinions are close enough. This reflects better the process of opinion sharing in conflicting situations. In order to narrow down our review, in the next section we select a model that is closest to our research requirements and compare it with the model discussed above. Before this, we focus on the main principle of the sharing process in social influence models to draw important conclusions that are relevant to our approach.

To describe the dynamics of opinion sharing in a society, Banerjee (1992) introduces the concept of information cascade, or herding. Its description is based on the following observation. Suppose that Bob wants to dine in a restaurant in an unfamiliar town, and based on his own research of the two available options, A and B, he chooses A. However, when he arrives at A, he sees that it is empty while B is crowded. He believes that other diners have similar tastes and that they may have some information about which is a better restaurant to eat at. Therefore, it may be rational for Bob to join the crowd at B rather than to follow his own information. Thus, he infers from the choices of others that his opinion, which is based on his own private information, might have been wrong. In this case, we say that an information cascade has occurred and that the following visitors are likely to make the same decision.

An information cascade has the potential to occur when agents make decisions sequentially, with later agents observing the opinions of earlier agents, from which they infer something. Ultimately, information cascades explain many types of imitation in social settings. Fashions and fads, voting for popular candidates or the popularity of a technological gadget can all be seen as examples of herding, in which people make decisions based on inferences from the actions or opinions of other people. Before describing an opinion sharing model based on this principle, we analyse a simple herding experiment created to illustrate how these models work (Anderson and Holt, 1996).

In particular, we assume the following settings: each agent has to make an opinion, for example, whether to adopt a new technology, eat in a new restaurant or support a particular political position; all agents form their opinions sequentially; and each agent can observe the opinions made by those who acted earlier. This model is simplified by the fact that the communication network is fully connected and the sequence of opinion formation is predetermined. Each agent has some private information that helps to guide its decision. However, the agent cannot directly observe the private information of others but it can infer this private information from their actions, which are based on their opinions. To build a mathematical model of this inference problem, it is assumed
that agents employ the Bayesian Theorem to determine the probabilities of events given observed information. Easley and Kleinberg (2010) made some observations about the outcomes of this experiment that are relevant to our study:

- Cascades can be wrong. If, for example, several agents initially had incorrect private information, a cascade of acceptances starts immediately, even though it is the wrong choice for the system.

- Cascades can be based on very little information, since agents rely more on new observations than on their own private information. This means that if a cascade starts relatively quickly in a large system, most of the private information that is collectively available to the individuals (in the form of their private beliefs) is not being used.

- Cascades are fragile. Since cascades can be based on relatively little information, they are easy to start; but this also makes them easy to collapse. If someone has a strong bias or receives slightly superior information that conflicts with the running cascade, it can overturn the cascade even if it has been running for a long period. Thus, the state of the system might be highly unstable.

The main lesson to be learned from studying cascades is to be careful in drawing conclusions about the best course of action from the behaviour of a crowd. As we have just seen, the crowd can be wrong even if everyone is rational and everyone has performed the same. The model we discuss in the next section addresses this problem by describing parameters that influence these cascading processes in networked societies.

### 2.3.1 Models with Dynamically Introduced Opinions

Against the general background on social dynamics we presented above, in this section we approach our research problem more closely. In particular, we select an appropriate model of opinion sharing that satisfies most of our research requirements, however it does not quantify the accuracy of consensus explicitly. Unlike the models discussed earlier, it enables us to reason about the accuracy of agents’ opinions by modelling the process of noisy introduction of new observations. In this section we briefly discuss the opinion sharing model developed by Glinton et al. (2009, 2010b,a, 2011) and then analyse properties of its dynamics when the opinions of agents are dramatically more accurate.

The aim of the model is to capture complex dynamics of opinion sharing about the true state $b$ of the common subject of interest $B = \{\text{orange, blue}\} \ (b \in B)$, in a large system of cooperative agents $A$ connected within a sparse communication network. In this model, some agents $S$, such that $S \subset A$ and $|S| \ll N$ have access to noisy sensors, and they introduce to the team conflicting opinions of which only one is correct.
The aim of each agent \( i \) and, as a result, of the whole system, is to form an opinion \( o_i \) that corresponds to the true state of the common subject of interest, \( o_i = b \). Following our discussion in Chapter 1, the frequency of finding this correct opinion over the number of simulations is the accuracy of the agent, and its averaged value over all agents defines the accuracy of consensus. However, in order to measure the performance of the system, Glinton et al. (2010a) proposed their own reliability metric as an average ratio between the total number of rounds, \( |M| \), each agent \( i \) has formed the correct opinion versus the incorrect one:

\[
R_{\text{ratio}} = \frac{1}{N} \sum_{i \in A} \frac{|\{m \in M : o^m_i = b\}|}{|\{m \in M : o^m_i \neq b\}| + 1}
\]  

(2.12)

where we add 1 to the denominator in order to avoid the undefined result. This definition implies that the team is heavily penalised for sharing the incorrect opinion by dividing by the number of agents that formed it. Therefore, \( R_{\text{ratio}} \) can be maximised even if the large share of the agents did not form their own opinions and their opinions stay undetermined. However, to date, this definition is the closest to measure the accuracy of consensus in modelling opinion sharing in large multi-agents systems.

A key assumption of the model is that, due to the communication constraint, agents can share with their neighbours only their new opinions without any additional information. Thus, each agent has to decide how informative the opinions it receives are, in order to form its own accurate opinion. Although restricting agents to communicating only their opinions is purely an abstraction to make working with and understanding the model easier, there are many real world domains where it is infeasible to communicate actual sensor readings. For example, the sensor data might be video or audio recordings that are expensive to share on a large network and might require significant effort and skills to interpret, or sensor data might be confidential or even consist of physical specimens that cannot be shared. If there are large numbers of sensor readings, restricted communication channels and many facts that a large number of agents need to come to conclusions about, we expect it to be infeasible to send most types of raw sensor data.

The process of decision making based on Bayes’ Theorem is similar to the herding model we discussed in the previous section. Each agent \( i \) uses either an observation received from a sensor or opinions about \( b \) communicated by its network neighbours, to form a private belief \( P_i(b = \text{orange}) \) about \( b \). A new observation \( o \) (let \( o = \text{orange} \)) is incorporated into the current belief to form a new belief \( P'_i(b = \text{orange}) \) using the following equation that is an expression of Bayes’ Rule:

\[
P'_i(b = \text{orange}) = \frac{w_i \cdot P_i(b = \text{orange})}{w_i \cdot P_i(b = \text{orange}) + (1 - w_i) \cdot (1 - P_i(b = \text{orange}))}
\]

(2.13)

where \( w_i = P_i(b = \text{orange} | o = \text{orange}) \) and \( w_i \) is the conditional probability that an observation \( o \) from a sensor reading or that a received opinion from a neighbour is correct. If the observation arrives from a sensor then \( w_i = r \) where \( r \) is the accuracy
of the sensor and an probability of how often on average the sensor observes the true state \( b \). We assume that sensors do not misreport, therefore the accuracy is limited to the range of \( 0.5 < r \ll 1 \). When an opinion arrives from any neighbour a weight \( w_i \) is assigned to it measuring the importance of this neighbour’s opinion. Since the agent does not have any additional information apart from the observation itself, it treats all observations as independent. Following our earlier discussion of the herding experiment, the treatment of observations by neighbours as independent is not correct, since they may have come to their conclusions based on the same data. Hence, agents relying on neighbours to form an opinion inevitably become over-confident in their conclusions due to double counting, where the same original observation is incorporated into the agent’s belief several times. However, without communicating actual sensor data or having detailed knowledge of the entire network structure and message sequence, it is impossible to completely remove this phenomenon.

In order for an agent to form its own opinion \( o_i \) based on its belief \( P_i(\text{orange}) \), the agent uses a simple threshold rule with confidence bounds, \((\sigma, 1 - \sigma)\); specifically if \( P_i(\text{orange}) \geq \sigma \), then \( o_i = \text{orange} \), and if \( P_i(\text{orange}) \leq 1 - \sigma \), then \( o_i = \text{blue} \). When the agent changes its opinion, it communicates its new value to all its neighbours. Subsequently, the neighbours may cross one of the confidence bounds and form an opinion cascade. Figure 2.5 illustrates a system in the process of opinion sharing, where some agents with sensors have already introduced opinions and cascades of conflicting opinions overlap. The probability \( P(c) \) that \( c \) agents change their opinions during the cascade is a key measure of the dynamics of the system and it was identified as an indicator of the performance.

Out of a number of opinion sharing models we discussed above, we select Glinton et al.’s model as a departure point to approach our the research problem. We support the well grounded critique that models of social dynamics are often too simplified to describe any real situation (Castellano et al., 2009). This is caused by a striking imbalance in the early work in the field of social dynamics between empirical evidence and theoretical modelisation, in favour of the latter. However, later developments in the field along with access to greater computing capabilities changed that perspective and have recently enabled scientists to model environments in greater detail.

In particular, Glinton et al.’s model is one of the most realistic and follows our motivating scenario discussed in Chapter 1. This model has a number of crucial properties that are not simultaneously present in the previous models of social dynamics, specifically:

- **Presence of a number of conflicting opinions in the team:** Many models describe the dynamics of social systems with a single type of information spread, whereas in our problem we deal with conflicting opinions introduced by sensors. This fundamentally changes the dynamics of the system and thus, we can cast aside
Figure 2.5: Sample system in the process of opinion sharing. Large red nodes are agents with sensors; smaller green nodes are agents that haven’t yet formed their opinions; white and black are agents that support corresponding opinions about the true state. (In the current round the correct state is orange. Grid topology is chosen for illustrative purposes, and irregularity in opinion cascades is caused by different prior beliefs of the agents.)

a large body of work, such as models describing propagation of fads (Bikhchandani et al., 1992), rumours (Nekovee et al., 2007) and gossip (Boyd et al., 2006).

- **Complex communication network**: Models of social dynamics often rely on a simplified representation of relations between individuals, such as homogeneous networks with regular or random topologies that can be analysed with the existing theoretical tools. This also applies to highly-acclaimed models introduced by social scientists, such as Schelling’s model for urban segregation (Schelling, 1971), Axelrod’s model for cultural dissemination (Axelrod, 1997) and a number of opinion sharing models, including the Voter model (Krapivsky, 1992; Frachebourg and Krapivsky, 1996), Sznajd model (Sznajd-Weron, 2000, 2005) and Majority rule model (Galam, 2002). However, as we discussed in Section 2.2.1 network topology has significant influence on the dynamics of sharing processes. Therefore, it is crucial to evaluate our approach on the complex networks we identified in Section 2.2.2.3.

- **Modelling of observations**: Widely recognised opinion sharing models (Voter model, Sznajd model, Majority rule model and others) assume that opinions are initially present in the system and focus on the analysis of their dissemination processes (Castellano et al., 2009). In contrast, in real-world settings, agents are often exposed to external factors that influence their opinions. The correct modelling of
this process of introducing new opinions into the system enables us to reason about the accuracy of the introduced information, and eventually about the accuracy of the agents’ opinions. This property of the model is crucial to meet our research aim of improving the accuracy of the agents’ opinions.

By combining these properties in a single model, we can argue that the solution developed for improving the model’s performance can be applied later in realistic scenarios. Apart from the abovementioned properties, Glinton et al.’s model is also the first model in which researchers discovered, and theoretically analysed, the influence of sharing dynamics on the accuracy of the agents’ opinions.

2.3.2 Opinion Cascades

In this section, we discuss the dynamics of the model and its influence on performance, specifically the accuracy of the agents’ opinions. In particular Glinton et al. (2010a) performed an analytical analysis of the model based on techniques from branching processes (Harris, 1963). They determined that the qualitative dynamics of the system are dependent on the value of the branching factor, which is the expected number of an agent’s network neighbours that change their opinions following the change of this agent’s opinion. It was discovered that in the state of scale-invariant dynamics, when the average branching factor is close to 1, the accuracy of consensus is dramatically improved. While the individual branching factors of agents may vary widely, creating an exponential distribution of cascade sizes, the average of 1 over the system leads to a balance between under- and over-estimating confidence in the propagated opinion that explains the improvement of accuracy of its consensus.

To describe the dynamics of the model, Glinton et al. developed a method to predict influence of the system parameters on $P(c)$, that is the probability that a cascade will encompass $c$ agents as a result of a single sensor observation. For this analysis, it is assumed that the network has a random topology with an infinite number of agents $N \to \infty$. These two assumptions taken in conjunction imply that there are no loops of neighbouring agents in the network. This allows the formulation of opinion cascades as a branching process parametrized by the branching factor, $\alpha$. For a given $\alpha$, which is the average number of an agent’s neighbours that adopt the same opinion on the next step, $P(c)$ follows directly from the theory of branching processes:

$$P(c) \propto c^{-3/2} e^{-\frac{\omega}{1+\alpha}}$$

(2.14)

where $\omega$ is a proportionality constant and $c$ is an independent variable. Thus, only $\alpha$ determines overall dynamics. Its value depends on the system parameters, specifically the agents’ weights, $w_i$, and the expected degree of the random network topology, $\langle d \rangle$. The branching factor, $\alpha$, is equivalent to the expected number of neighbours that change
Chapter 2 Related Work

Figure 2.6: Required number of belief updates \( n_\alpha(w_i) \) dependent on a weight \( w_i \) to form an opinion given an opposing belief \( P_i \) to this opinion

their beliefs when a random agent changes its belief. Following analysis using the mean field assumption, that has roots in statistical physics as discussed earlier, Glinton et al. identified the dependency of \( \alpha \) on the system parameters assuming that a weight \( w_i \) is common for all agents:

\[
\alpha = \frac{\langle d \rangle}{n_\alpha(w_i) + 2}
\]  

(2.15)

where \( n_\alpha(w_i) \) is the number of sequential observations, having the same value, that would be required to change the opinion of an agent starting with a belief that opposes those observations. We can calculate \( n_\alpha(w_i) \) by inputting different values of \( w_i \) into the belief update rule in Equation 2.13 and find the number of observations required to move a prior belief from either end of the belief range to the other end of the range. Following the intuition that with a higher weight an agent requires fewer updates of its belief to form an opinion. The resulting plot of \( n_\alpha(w_i) \) is shown in Figure 2.6.

This analysis of the branching factor \( \alpha \) identified three cases of distinct qualitative dynamics in the model, each resulting in drastically different performance. Following Equation 2.15 and Figure 2.6 it is possible to choose values of \( w_i \) and \( \langle d \rangle \) that result in different \( \alpha \) values. Specifically, there are three main cases:

- **Scale-Invariant Dynamics** – when parameters \( w_i \) and \( \langle d \rangle \) are chosen such that \( \alpha = 1 \). When this condition is satisfied Equation 2.14 reduces to:

\[
P(c) \propto c^{-3/2}
\]  

(2.16)

A probability distribution with this characteristic is traditionally known as a scale invariant distribution. Glinton et al. (2009) argued that scale-invariant dynamics corresponds to the phase transition in the opinion sharing process between the following two states observed on a wide range of parameters:
- **Stable Dynamics** – when $\alpha < 1$ then Equation 2.14 reduces to:

$$P(c) \propto e^{-3/2} e^{-\frac{c}{1-(\langle d \rangle / (n_o(w_i)+2))}}.$$  \hspace{1cm} (2.17)

Here the exponential factor has a negative sign, which means that the probability of larger cascades relative to the system size drops dramatically. In contrast to the scale invariant dynamics where cascades of all sizes are probable, in this case cascades quickly decay after the sensor reading.

- **Unstable Dynamics** – when $\alpha > 1$ then Equation 2.14 reduces to Equation 2.17, however the sign on the exponential term becomes positive. Consequently, this results in frequent large cascades.

We reproduce the results published by Glinton et al. (2010a) for illustrative purposes. Specifically, we simulate the model of $N = 1000$ agents with $|S| = 50$ sensors and a random network with average degree $\langle d \rangle = 8$, where results are averaged over 100 opinion dissemination rounds. In this case, the theoretically predicted critical weight, when scale-invariant dynamics are observed, is $w_{\text{critical}} \in [0.63..0.64]$ (for $\langle d \rangle = 8$ according to Equation 2.15 $n_o(w_i) = 6$, and is shown with a dash line in Figure 2.6 for this case $w_{\text{critical}} \in [0.63..0.64]$). Figure 2.7 confirms this and that the accuracy of consensus is maximised in this area of parameters. The analysis of dynamics in this area shown in Figure 2.8, confirms that we observe scale-invariant dynamics.

![Figure 2.7: The state of the agents at the end of simulation and the accuracy of consensus depending on a value of weights. The reliability metric is maximised when weights are close to the theoretically predicted, $w_{\text{critical}} = 0.64$.](image-url)

The area of improved accuracy of consensus can be explained in terms of model’s dynamics. The frequent smaller cascades prevent the system from overreacting to incorrect
opinions, however, though less frequent, large cascades can occur and disseminate these locally-supported opinions to the rest of the agents. This hypothesis is supported by the results presented in Figure 2.9. It shows a scatter plot of the size of cascades against the average belief of the system in the previous belief update step: \( \frac{1}{N} \sum_{i \in A} P_i(b = \text{orange}) \) (assuming that \text{orange} is the correct opinion).

However, as might be expected after our discussion of different network structures, Glinton et al. (2010a) showed that these theoretical results are suitable only for random networks and cannot be generalised for complex network topologies. Therefore, there is a clear need to develop a solution that reaches the area of these optimised parameters for any complex network topology, as we stated in our research aim. In the following section we review Glinton et al.’s existing algorithm, DACOR, which attempts to fill this gap. Specifically, DACOR exploits the properties of scale-invariant dynamics and tunes the system parameters in order to improve the accuracy of the agents’ opinions.

### 2.4 Benchmark Algorithm for Achieving Accurate Consensus

To reach this area of optimised performance in a complex communication network, Glinton et al. (2010a) proposed the Distributed Adaptive Communication for Overall Reliability (DACOR) algorithm. DACOR adjusts the agents’ weights according to the estimated local branching factor, \( \alpha_i \). As mentioned earlier, it is impossible to predict weights that induce the described emergent behaviour. Therefore, the algorithm gradually improves the accuracy of consensus by tuning weights through a number of opinion dissemination rounds.
Figure 2.9: The size of opinion cascade depending on the average belief. Different cascades are marked as following: ‘correct’ ‘+’, ‘incorrect’ ‘x’. When \( w_i = 0.6 < w_{\text{critical}} \) opinions do not disseminate on large scale, and when \( w_i = 0.66 > w_{\text{critical}} \) we observe large cascades of incorrect opinion that reduces the accuracy of consensus. In the area close to the critical value \( w_i \in w_{\text{critical}} = [0.63, 0.66] \) large cascades share only correct opinion.

DACOR is based upon the observation that the accuracy of consensus of the model is maximised when the branching factor is close to 1, when scale-invariant dynamics are observed. Since each agent \( a_i \) can observe how many neighbours have changed their opinions at each step of their belief updates \( c_i = |\sigma_j^k \neq \sigma_j^{k-1} : j \in D_i| \), it can estimate the local branching factor \( \alpha_i \). Following this, it communicates to all its neighbours how much \( \alpha_i \) deviates from 1 in an attempt to cooperatively achieve \( \alpha_i = 1 \). In order to do so, all agents that received a message with \( \Delta \alpha_i \), that is the difference between 1 and actual \( \alpha_i \) in the neighbourhood \( D_i \), adjust their weights to compensate for this.

In more detail, Algorithm 2.4 presents the pseudo-code of DACOR separated into two corresponding procedures: SendMessage, that is executed when an agent observes opinion changes in its neighbourhood; and ReceiveMessage, that corrects the weight of the agent when it receives a service message. In line 1 of SendMessage the agent
calculates its local branching factor, $\alpha_i$, where $c_i$ is the number of the agent’s neighbours that have just changed their beliefs and $u$ is a factor that gives more weight to recent local observations of $\alpha_i$. Then in lines 2-5 the agent sends its approximation of the local $\Delta \alpha_i = \alpha_i - 1$ to its neighbours $j \in D_i$. When the agent receives such messages, it executes RECEIVEMESSAGE and in line 1 it updates its weight, $w_i$, proportionally to $\Delta \alpha_i$ and its derivative $\Delta \alpha_i'$ to compensate for oscillations. The remaining lines ensure that the weight, $w_i$, remains in the range $[0.5, 1]$.

Algorithm 1 DACOR\(^1\): Benchmark Algorithm for Achieving Accurate Consensus

**Procedure** SENDMESSAGE($i, c_i, u = 10$)

**Require:** $c_i \geq 1$

1: $\alpha_i = \alpha_i(u - 1)/u + c_i/u$

2: $\Delta \alpha_i = \alpha_i - 1$

3: for all $j \in D_i$ do

4: RECEIVEMESSAGE($a_j, \Delta \alpha_i$)

5: end for

**Procedure** RECEIVEMESSAGE($i, \Delta \alpha, \gamma = \frac{1}{1000}, \beta = \frac{1}{10}$)

1: $w_i = w_i - \gamma \Delta \alpha + \beta(\Delta \alpha_i' - \Delta \alpha)$

2: $w_i = \text{LIMITWITHRANGE}(w_i, [0.5, 1])$

3: $\Delta \alpha_i' = \Delta \alpha$

The procedure SENDMESSAGE is executed by every agent in the network that has observed new opinions in the last step of belief update ($c_i \geq 1$). After receiving a message, an agent executes the procedure RECEIVEMESSAGE that updates its weights. Thus, if an agent changes its opinion, all its neighbours communicate on average $\langle d \rangle^2$ additional messages to tune the weights, where $\langle d \rangle$ is the average degree of the network. Therefore, actually performing a decentralised estimation of the branching factor requires significant message overhead compared to the number of messages used to share opinions. Additionally, as our empirical evaluation reveals, the internal parameters of DACOR are sensitive to the system’s configuration and DACOR has to be tuned individually for different domains. We address these shortcomings by presenting novel approaches in following chapters.

2.5 Summary

In this chapter we reviewed the scientific literature relevant to our work. We started by discussing how consensus is reached in large systems, its properties and defining our goal of maximising its accuracy. Then, we focused on modelling such systems using the multi-agent paradigm. We started from analysis of the structure of decentralised systems, specifically the topological properties of their communication networks in the light

\(^1\)The DACOR algorithm published by Glinton et al. (2010a) contains a misprint in the calculation of $\alpha_i$. With the help of the authors we describe the corrected implementation in Algorithm 1.
of their influence on the information sharing process. Considering the high interdependence between different topological properties, we selected three network generators with distinct properties that are most discussed in the literature. This enables the evaluation of the adaptivity and robustness of our solutions in the following chapters.

Next, we briefly examined the literature on existing models of sharing processes in large teams. After reviewing the research on social dynamics, we explained cascading behaviours in teams and their implications. Given this, we introduced our problem of improving the accuracy of consensus by exploiting the properties of sharing processes.

To find a suitable model of the environment we briefly reviewed existing opinion sharing models and chose Glinton et al.’s model as a departure point. Importantly, Glinton et al. have recently discovered properties of dynamics in this model that indicate a state in which the agents’ opinions become dramatically more accurate. Moreover, we identified that exploiting these properties is the most efficient approach to improve the accuracy of consensus in the restricted settings defined by our research requirements.

However, following our discussion of the influence of the topology of the communication networks on dynamic processes in multi-agent systems, it is apparent that it is extremely difficult to predict the system parameters when the accuracy of consensus is maximised. Therefore, Glinton et al. (2010a) proposed the adaptive algorithm, DACOR, for reaching these parameters in a decentralised fashion. However, in order to operate, DACOR requires significant communication overhead to exchange service messages compared to the communication required to share the opinions. This violates our research requirement of minimal communication, and cannot be used in the settings of our motivating scenario. Against this background, we present in the next chapters our model of the environment and algorithms to improve the accuracy of consensus in a decentralised fashion.
Chapter 3

Modelling Collective Behaviour in Opinion Sharing

To approach our research aim of improving the accuracy of consensus in large multi-agent systems, we need to formalise a model of such an environment, its agents, and the opinion sharing processes between them. In the literature review we discussed a number of opinion sharing models that formalise this problem. However, we showed that none of these models match our motivating scenarios. Therefore, in this chapter we present a new opinion sharing model.

In designing our model in Section 3.1 we build upon the most promising model offered by Glinton et al. (see Section 2.3.1). To address the shortcomings we identified in this model, we introduce a number of crucial modifications. Specifically, our opinion sharing model is the first to measure the specific impact of collective behaviour on the accuracy of consensus, which is an essential requirement to approach our problem. In order to do this, we change the agent model and choose new metrics that are more closely aligned to our research requirements. Following this, we analyse the theoretical bounds on the performance metrics.

In our search of system parameters that lead to an improved accuracy of consensus, in Section 3.2 we choose experimental setups to cover a wide range of model parameters. Our analysis of these settings in Sections 3.3 and 3.4 reveal common patterns. We rely on this analysis in designing our decentralised algorithms for adaptive accuracy improvement, which we present in the following chapters. Most importantly, we show that Glinton et al.’s analysis suggesting that the value of the branching factor indicates the state with the highest performance does not hold in our model. We conclude this chapter with a number of benchmarks presented in Section 3.5. Their purpose is to show the level of accuracy improvement that can be achieved when the properties and parameters of the model are known.
3.1 Opinion Sharing Model

The proposed agent-based model of opinion sharing is a generalisation of the existing models with an emphasis on the process of introducing new observations into the system. This focus enables us to approach our research problem and analyse the exact connections between the accuracy of observations and the accuracy of consensus. Specifically, unlike the classical models of social dynamics where agents are initially endowed with opinions (see Section 2.2.3), our model is built upon Glinton et al.’s recently-offered model in which new opinions are introduced gradually. It was found that this additional level of detail in modelling changes the dynamics of the opinion sharing process, which is crucial for the accurate representation of realistic settings.

However, as discussed in Section 2.3.1, the latter model has a crucial shortcoming which prevents analysis of the exact impact of the collective behaviour on the accuracy of consensus. Specifically, its process of introducing new opinions requires that agents aggregate a number of observations before forming their own opinions, thus implementing a form of local filtering. This design implies that: (i) agents with sensors may never form their opinions if they do not receive enough observations; (ii) speed of convergence to the consensus cannot be measured, since sharing of the observations is delayed until the sensing agents are confident enough to form and share their opinions; (iii) improvement of the accuracy of consensus is a combination of collective behaviour, and a particular design of local filtering procedure.

In contrast, in our model presented in the following Section 3.1.1 new opinions are introduced as direct changes of agents’ opinions. Crucially, this simplification of external influence allows us to focus only on the impact of collective behaviour on the system performance, ignoring specific design of the observation process. Additionally, in order to avoid misleading conclusions which might be caused by our choices in agent design, we offer two alternative agent designs. Later, a comparative analysis of these two designs in Section 3.3 reveals which metrics are more reliable indicators of the critical parameters that lead to the highest accuracy of consensus. In the following chapters this enables us to evaluate the adaptivity of our decentralised algorithms in finding these critical parameters. In order to provide an extensive evaluation, we offer in Section 3.1.2 a number of metrics that follow our research aims. Finally, we conclude the model description in Section 3.1.3 by comparing it with the existing models.

3.1.1 Model Description

Formally, the model is defined on a communication network which is a connected graph with a certain degree distribution. The nodes of the network represent agents, denoted as \( A = \{i^1 \ldots i^N\} \) where \( N \) is the number of agents in the system, and the edges of the network indicate which agents are neighbours and can therefore communicate. The
aim of each agent $i \in A$ is to form its own opinion, $o_i$, such that it matches the correct opinion, $b$, which describes the true state of the subject of common interest, $B$, where $b \in B$. In our examples we use $B = \{\text{orange, blue}\}$. For illustrative purposes we assume that the correct opinion is $b = \text{orange}$. Following the discussion of the existing models in the Chapter 2, we also support an assumption that $B$ can be limited to a binary set. This assumption follows the argument that a binary choice can be applied to a wide range of real world situations (Watts and Dodds, 2007). In designing our decentralised solutions in the following chapters, we do not rely on this assumption to generalise our findings. However we incorporate it into our model in order to simplify its notation.

Despite the fact that the common subject of interest is binary, initially all agents hold an undetermined opinion, thus $o_i \in B \cup \{\text{undetermined}\}$. In order to recover the correct opinion an agent relies on the opinions of their network neighbours, $D_i = \{j^1 \ldots j^d_i\}$ where $d_i$ is the number of neighbours. This restriction on communication paths clearly indicates the influence of the topology of the underlying communication network on the opinion sharing processes. New opinions are introduced into the system only by a small subset of $N_s$ sensing agents, $S \subset A$, $N_s \ll N$, which are the event witnesses in our motivating scenarios. These sensing agents may form their opinions not only by relying on their neighbours, but also under an external influence which corresponds to an observation of the subject of common interest. We model this as a direct change of an opinion of a randomly selected sensing agent, $i \in S$. In order to compare convergence of the system in different settings, new opinions are introduced with a constant rate, $\lambda$, which is the number of opinion update steps between the introduction of new opinions.

Crucially, the new opinions have low accuracy, $r$, which is the probability of the opinion being correct. Similarly to the existing opinion sharing models, we assume that the agents are cooperative and non-malicious. Therefore, the accuracy of the sensing agents is limited to the range of $0.5 < r \ll 1$ (or in percentile terms $50\% < r \ll 100\%$). This implies that the number of correct opinions introduced into the system is slightly higher than the number of incorrect ones. Following every change of its opinion, each agent communicates it to all of its neighbours participating an opinion sharing cascade started by a sensing agent.

In order to decide which opinion to adopt, the receiving agent, $i$, updates its private belief, $p_i$, by starting from a prior belief, $p_i^0$, and applying a decision rule to form its own conclusion on which opinion to support. In the process of its belief update, agent $i$ applies an aggregation function to new opinions received from its neighbours $o_j$, $j \in D_i$ with a certain weight attributed to each neighbour $w_{ij}$:

$$p_i^k = f \left( p_i^{k-1}, o_j, w_{ij} \right)$$  \hspace{1cm} (3.1)

where $k$ is a belief update step. The weight $w_{ij}$ represents the importance of a received opinion and it encodes the social influence of agent $j$ on $i$. The key aim of agent $i$ is to find the critical values of its weights $W_i = \{w_{ij} : j \in D_i\}$ such that it will be forming
the correct opinion, which is never observed directly. In our model we assume that set of weights is the only parameter that an agent has influence on. Figure 3.1 illustrates a sample structure of such a multi-agent system with an emphasis on the fact that weights might not be symmetrical or equal.

The behaviour of the whole model, such as we provided in the sample above, depends on the activities of individual agents. Figure 3.2 explains the model of a single agent and we will now discuss its components in more detail. To show later that the properties of the model do not depend on a specific model of an agent, we consider two agent designs. The first agent design is inspired by Glinton et al.’s agent design, whilst the second one is widely used in the previous opinion sharing models. Their principal difference lies in their aggregation function, $f \left( p_i^{k-1}, o_j, w_{ij} \right)$:

1. **Bayesian aggregation function** which is based on Bayes’ theorem. In our case, in which the subject of common interest is binary, we can assume that $p_i^k$ is the probability that $b = \text{orange}$ and consequently $1 - p_i^k$ is the probability of $b = \text{blue}$. Following this, the aggregation function can be defined using Bayesian updating as:

   $p_i^k = \frac{w p_i^{k-1}}{(1 - w)(1 - p_i^{k-1}) + w p_i^{k-1}}, \quad (3.2)$

   where

   $w = w_{ij}$ \quad if $o_j = \text{orange}$

   $w = 1 - w_{ij}$ \quad if $o_j = \text{blue}$

   where $w_{ij}$ is a conditional probability that agent $j$ reports the correct opinion.

   In this case $w_{ij} = 0.5$ indicates that the received opinion is ignored, and with $w_{ij} = 1$ the agent changes its belief to $p_i^k = \{1, 0\}$, depending on the received
Chapter 3 Modelling Collective Behaviour in Opinion Sharing

51

Aggregation function

\[ p_i^k = f(p_{i}^{k-1}, o_j, w_{ij}) \]

Prior

\[ p_{i}^{0} = p_{i}' \]

Decision rule

\[ o_i^k = F(o_{i}^{k-1}, p_{i}^k, \sigma) \]

Opinion \( o_i^k \)

has opinion changed?

\[ o_i^k \neq o_{i}^{k-1} \]

Yes

Agent \( i \)

Figure 3.2: Model of an agent

opinion regardless of its previous value. In our model we assume that agents are cooperative and have no intention to misreport, therefore we limit \( w_{ij} \) to the range of \([0.5, 1]\). This rule was offered earlier and enables us to provide a direct comparison with the existing results (Glinton et al., 2009; Pryymak et al., 2012). Figure 3.3a illustrates an aggregation process in which an agent first receives 4 opinions of \( o_j^k = \text{blue} \) followed by 11 contradictory opinions.

2. Weighted sum aggregation function. This type of aggregation function was initially proposed by DeGroot (1974) for opinion sharing models and is also known as an
Chapter 3 Modelling Collective Behaviour in Opinion Sharing

Figure 3.3: Agent’s aggregation functions and moments of opinion formation following the threshold decision rule. (Assuming that agent \( i \) attributes the same weight to all of its neighbours.)

imitation rule. It is given by:

\[
p_k^i = p_{i}^{k-1} + w, \tag{3.3}
\]

where

\[
\begin{align*}
  w &= 2(w_{ij} - 0.5) \quad \text{if } o_j = \text{orange} \\
  w &= -2(w_{ij} - 0.5) \quad \text{if } o_j = \text{blue}
\end{align*}
\]

where \( w \) is scaled in order to keep the same search range \([0.5, 1]\) of agents’ weights as in the previous aggregation function. This additional aggregation rule is used as a benchmark in order to identify which metrics are influenced by a specific agent design and which are more universal indicators on the state with the highest accuracy of consensus. Figure 3.3b illustrates its sample dynamics and shows the principal difference to the Bayesian aggregation.

These aggregation functions are used by the agent to update its private belief with a number of opinions received from its neighbours. In the next step, the agent has to decide whether it is confident enough to form its own opinion, \( o_k^i \). In order to do so, the agent applies a decision rule to its private belief. In our model we adopt a widely
studied threshold rule (Watts and Dodds, 2007), which is a sharp hysteresis function:

$$a^k_i = F\left(a^{k-1}_i, p^k_i, \sigma\right) = \begin{cases} 
\text{undeter.} & \text{initial, if } k=0, \\
\text{orange} & \text{if } p^k_i \geq \sigma, \\
\text{blue} & \text{if } p^k_i \leq 1 - \sigma, \\
\text{other}\phantom{s} & \text{otherwise}
\end{cases}$$

(3.4)

where thresholds \(\{1 - \sigma, \sigma\}\), \(\sigma \in (0.5, 1)\) are the confidence bounds upon crossing which the agent changes its opinion. The shape of this function is shown in Figure 3.4, and the corresponding moments of opinion changes were illustrated earlier in Figure 3.3.

Every time the agent changes its opinion, it communicates the new opinion to all its neighbours. Consequently, these neighbours update their own beliefs and may form their own new opinions. If the agent changes its opinion following a received opinion from its neighbour, it participates in an opinion cascade where a number of agents change their opinions in a sequence after a critical new opinion. Figure 3.5 illustrates the sample dynamics in the model. Here the plot shows rapid changes in the number of agents supporting each opinion and the network reflects the state of the system indicated on the plot. The complex pattern of agents supporting different opinions resulted in a number of opinion cascades.

Note, that since the agents form their opinions based on their private beliefs, the introduction of new opinions are implemented as direct changes of their beliefs to the values that are minimally required to form the corresponding opinion:

$$p^k_i = \psi\left(\lambda, r, b, \{\sigma, 1 - \sigma\}\right)$$

(3.5)

where \(\{\sigma, 1 - \sigma\}\) are the corresponding confidence bounds of the decision rule, \(\lambda\) is the above-mentioned rate of introducing new opinions, and \(r\) is the accuracy of introduced opinions, which is the probability of a new opinion corresponding to the correct state \(b\).
Chapter 3 Modelling Collective Behaviour in Opinion Sharing

3.1.2 Performance Metrics

In order to accurately measure the performance of the system, we define all performance metrics as their average values over a number of opinion sharing rounds, \( m \in M \). In order to evaluate the convergence of the system later, we assume that each round, \( m \), is limited by a fixed number of sensor observations, \( A \). Therefore, eventually the processes of opinion sharing stop and this corresponds to the end of a round. We assume that this constitutes a deadline when the subject of common interest may be changed, or its correct state may get a new value. Thus, in the beginning of each new round we reinitialise all agents with the undetermined opinion, their beliefs with the original priors, and restart the sharing process with a new, randomly-selected, correct state \( b^m \in B \).

Following the research aims introduced in Section 1.4, we next define the accuracy metric along with its theoretical bounds. To study compliance with the remaining the research requirements, we also formalise the metrics to quantify the communication expense in Section 3.1.2.2 and the convergence to the consensus in Section 3.1.2.3.

Figure 3.5: Sample dynamics of the model during an opinion sharing round. Irregularities in opinion cascades are caused by different prior beliefs and different weights agents attribute to their neighbours.
3.1.2.1 Accuracy of Consensus

The accuracy of consensus is the most important metric for this research and its maximisation corresponds to the main aim of this work. Following our motivating scenarios and discussion of the aims in Chapter 1, we offer a metric which is maximised when all agents form the correct opinion. Formally, our accuracy metric measures how often an agent is expected to form the correct opinion over a number of opinion sharing rounds, $|M|$: 

$$R = \frac{1}{N|M|} \sum_{i \in A} |\{m \in M : o_{imb}^i = b^m\}| \cdot 100\% \quad (3.6)$$

where we count the number of opinion sharing rounds after which each agent has formed the correct opinion, $b^m$, and normalise by the number of agents and sharing rounds.

In order to evaluate our model and methods of improving the accuracy of consensus, it is important to establish indicative bounds on this metric. Specifically, our approach enables us to determine analytically the minimum and maximum levels of the accuracy metric. In contrast to Glinton et al.’s model, we are able to do so since in our model new opinions are introduced directly into the system thereby avoiding the local filtering of observations on an agent level. In our later analysis these accuracy bounds will indicate the relative performance which can achieved by tuning the system in comparison to the theoretical maximum and minimum.

In particular, the most accurate opinion can be formed by directly aggregating all opinions introduced into the system. However, this would require a central authority which aggregates opinions from all sensing agents, makes its decision on which opinion is correct and shares it with the rest of the agents. In terms of our model, such settings are observed when agents form a star topology (as shown in Figure 3.6). Following our review of the accuracy of consensus in a centralised system, especially the notion of ‘group intelligence’ in Section 2.1.1, we can apply Condorcet’s jury theorem in order to derive the upper boundary of the accuracy metric in such idealistic settings. Here we repeat its derivation applied to our model.

In order to form the most accurate opinion in such a scenario, the central authority, or simply the centre, must follow the majority rule and form an opinion which is simultaneously supported by at least $\lceil N_s/2 \rceil$ sensing agents. Since we assume that the subject of common interest is binary, the opinions received by the centre follow a Bernoulli trial. This enables us to calculate the accuracy of the centre, since the accuracy of opinions reported by the sensing agents, $r$, is known. Therefore, the accuracy of the centre is the probability that more than half of the sensing agents report the correct
Chapter 3 Modelling Collective Behaviour in Opinion Sharing

opinion:

\[ r_{\text{centre}} = Pr(K > N_s/2) \]  

(3.7)

where \( K \sim B(N_s, r) \) is the binomial distribution which describes the received binary opinions, and \( Pr(K > N_s/2) \) is its cumulative distribution function. Thus, it can be unfolded as a sum of probabilities of all cases when more than half the sensors report the same opinion:

\[ r_{\text{centre}} = \sum_{i=\lceil N_s/2 \rceil}^{N_s} \binom{N_s}{i} r^i (1 - r)^{N_s - i} \]  

(3.8)

To illustrate our results, assume that \( N = 1000 \) is the number of agents in a system which includes \( N_s = 0.05 \cdot N = 50 \) sensing agents. The accuracy of a new opinion each sensing agent can observe, which is the probability of the correct opinion, is a fixed value, \( r = 65% = 0.65 \). In this case \( r_{\text{centre}} = 98.99\% = 0.9899 \). Knowing the accuracy of the centre, we can calculate the maximum of the accuracy of consensus following its definition in Equation 3.6:

\[ R_{\text{max}} = r_{\text{centre}} \cdot \left( 1 - \frac{(1 - r)N_s}{N} \right) \cdot 100\% = 97.26\% \]  

(3.9)

where the fraction represents a share of the sensing agents that are expected to form the incorrect opinion. Our empirical evaluation confirmed this figure: \( R_{\text{max}} = 96.09 \pm 1.84\% \).

Now, our motivating scenarios focused on networked systems which have complex topologies, and thus, are unlikely to use centralised decision making. For such cases we need to look at the worst case scenarios and analyse the minimum level of accuracy that can be achieved. Specifically, we consider two scenarios and corresponding definitions of the minimal accuracy of consensus:

- When opinions are not shared in the system and only sensing agents form their opinions (which corresponds to the stable state of the system dynamics):

\[ R_{\text{min1}} = \frac{rN_s}{N} \cdot 100\% = 3.25\% \]  

(3.10)

- When an opinion from a single sensing agent is adopted by all agents, and thus, they do not benefit from the presence of several opinion sources in the system (which corresponds to the unstable state of the system dynamics):

\[ R_{\text{min2}} = r \cdot \left( 1 - \frac{(1 - r)N_s}{N} \right) \cdot 100\% = 63.86\% \]  

(3.11)

where similarly to Equation 3.9 the fraction represents a share of the sensing agents that are expected to form the incorrect opinion.

Building on this, Figure 3.7 shows how the bounds on the accuracy of consensus scale with the size of the system. As can be seen, the maximum accuracy, \( R_{\text{max}} \), fluctuates as
the number of sensing agents, \( N_s = 0.05 \cdot N \), changes from an even to an odd number, affecting the accuracy of the centre, \( r_{\text{centre}} \). The behaviour of this upper boundary suggests that the accuracy of consensus is expected to rise with the size of the system, approaching \( \lim_{N \to \infty} r_{\text{centre}} = 1 \), \( R_{\text{max}} \to 100\% \) for a system with a star topology communication network of infinite size. At the same time, the lower boundary, \( R_{\text{min2}} \) indicates the minimal level which should be achieved by any methods of improving the accuracy of consensus.

The second important metric, after the accuracy of consensus, is designed to verify the compliance with our key restriction to the environment, which is minimal communication in the system.

### 3.1.2.2 Communication Expense

Following our research requirements, we have already restricted the communication in the model to that of opinion sharing, specifically, by prohibiting the sharing of any additional information other than the state of an agent’s opinion (see discussion in Section 3.1.1). With this metric, we study how much communication remains and the minimal amount required for agents in the model to form their opinions.

In particular, we define the communication expense, \( U \), as the number of messages that are transmitted in the system during an opinion sharing round. Each message carries an opinion without any supporting information from a sender to a single recipient. In order to define the minimal communication in the system we rely on the fact that to maximise the accuracy \( R \), all agents have to form their opinion. Thus, each agent has to share its opinion at least once and the minimal number of messages required to share
an opinion is:

\[ U_{\text{min}} = N \cdot \langle d \rangle \] (3.12)

where \( N \) is a total number of agents and \( \langle d \rangle \) is the expected number of neighbours of each agent. This number of messages, \( U_{\text{min}} \), is necessary to share an opinion from a sensing agent to the rest of the system in a single opinion cascade.

In the following evaluation of our model, we analyse the communication expense by measuring its expected value over a number of independent opinion sharing rounds, similar to what we offered for the accuracy metric. In designing online solutions for accuracy improvement in the following chapters, the lower boundary on communication, \( U_{\text{min}} \), will become an important indicator on the relative expenses required by the solutions offered.

Finally, the last of the performance metrics is designed to measure the timeliness of agents’ opinion formation, in addition to their accuracy and communication expense.

### 3.1.2.3 Convergence to Consensus

Another important performance metric is the *convergence* to consensus. Our model converges in a number of sudden steps which correspond to the occurrence of large opinion cascades (see example in Figure 3.5). As the result of this, the dynamic rate of convergence is not constant and cannot be used as an indicative metric. Therefore, we define our convergence metric, \( C \), as an expected opinion update step, \( k \), when at least 80% of the agents form the same opinion for the first time (the value of this threshold is adopted from Glinton et al. (2010b)). This metric measures the timeliness of the opinion formation and does not take in account whether or not the consensus is correct. Moreover, we consider unanimity to be a very rare event, and thus, apply a majority definition of consensus which does not require all agents to form the same opinion (in line with previous work in this area, see Section 2.2.3).

In order to avoid distortion of the average value of the convergence metric, we exclude opinion sharing rounds when the team did not reach the threshold level of accuracy. Similar to defining the minimal communication expense, the minimal convergence is the number of update steps required to share an opinion in a single cascade. This number of steps depends on the topology of the communication network and corresponds to its diameter, which is the longest of all the calculated shortest paths in a network:

\[ C \propto \max (l) \] (3.13)

However, our model contains a number of sensing agents which can introduce observations simultaneously. Therefore, its more appropriate to choose as the benchmark the
average shortest path length:

$$C_{\text{min}} \approx \langle l \rangle$$ (3.14)

where $\langle l \rangle$ is defined in Equation 2.5 in Section 2.2.1.

Having introduced the metrics, we now discuss how our model differs from the existing opinion sharing models.

### 3.1.3 Comparison Against the Existing Models

In our literature review in Section 2.2.3 we discussed in detail the existing opinion sharing models and concluded that in order to approach our research problem, a new model is required. Here we briefly repeat our motivation and provide an overview of the principal differences in our model design compared to the existing models:

- **Presence of conflicting opinions.** In the opinion sharing models based on the Ising magnetism and epidemic models, agents can only share a single type of information (rumour or infection) without a contrary type. In contrast, in our model, opinions may be conflicting and each agent has to make a decision as to which one to support. This is the first step to enable reasoning about the accuracy of formed opinions.

- **Gradual introduction of new opinions.** In the classical opinion sharing models developed from the Ising magnetism model, such as Voter and Sznajd models, agents are initially endowed with opinions. However, it is more realistic for many scenarios to allow the gradual introduction of new opinions, assuming that agents are neutral at the beginning. Crucially, this dramatically changes the system dynamics. As we show later, the opinion cascades, which are initialised as a result of the process of gradual introduction, create specific circumstances in which distributed opinion aggregation is possible. Thus, our problem of improving the accuracy of consensus can be approached by exploring these properties.

- **Notion of accuracy of the opinions and consensus.** Specifically, we assume that one of the conflicting opinions is more common amongst those introduced into the system. This opinion corresponds to the correct state and the level of its domination is the accuracy of the observations. We explicitly define the accuracy of consensus in a similar way. Previously, the closest model to ours, the model developed by Glinton et al., was analysed from the perspective of its reliability metric, which is defined by other units of measurement than accuracy of opinion. As a result, their reliability metric is not maximised when most of the agents form the correct opinion.

- **Quantification of the accuracy improvement due to the collective behaviour.** Unlike in Glinton et al.’s model, in our design we avoid local filtering
of observations on an agent level, and introduce new opinions into the system directly. This enables us to quantify improvement of the accuracy of consensus, and derive analytical bounds on the performance metrics.

- **Alternative agent designs.** In order to abstract our study of the model dynamics from its implementation details, we offer two different variations of the agent opinion formation process. By doing so, we can avoid some of the model bias and make more generic conclusions on the factors that are the most influential on the accuracy of consensus.

The first two features differentiate our model from most of the existing models. The last one stands out in the analysis of the model’s dynamics, and it will later enable us to verify if our techniques are sensitive to specific agent design. While this feature is discussed in detail in the analysis of the computational results, the rest of the differences compared to Glinton et al.’s model require additional overview.

As we discussed in Section 2.3.2, the main analysis used in Glinton et al.’s work was in terms of a reliability metric, $R_{\text{ratio}}$, which is defined as the average ratio (over all agents) between the number of opinion rounds when the correct opinion was formed and the number of rounds with an incorrect opinion (see Equation 2.12). Such a definition implies that it is better for an agent not to form an opinion and stay undetermined since the metric heavily penalises incorrect opinions. In contrast, our accuracy metric, $R$, is maximised when the agents form the correct opinion as often as possible, despite occasionally forming an incorrect opinion. Moreover, our metric is defined on the same scale as the accuracy of new opinions introduced. This enables us to analyse the level of accuracy improvement that can be achieved by exploiting the properties of our model. Finally, our metric follows our motivating scenarios, which assume that opinions provide important information for agent activities and staying undetermined is close to being incorrect. Therefore, we consider $R$ as a more plausible metric for measuring performance in our case.

A direct empirical comparison of our opinion sharing model presented above (in red) and Glinton et al.’s model (in blue) along with the reliability metric, $R_{\text{ratio}}$, and our accuracy metric, $R$, is shown in Figure 3.8. In this experiment we used the same settings in which Glinton et al.’s model was evaluated earlier in Section 2.3.2. Specifically, in this experiment we evaluate both models on a number of scale-free networks of $N = 1000$ agents with the Bayesian aggregation function which is common to both models (the rest of parameters are: $\langle d \rangle = 8$, $r = 65\%$, $\sigma = 0.8$, $p_i' \in \mathcal{N}(\mu = 0.5, s = 0.1)$, $|M| = 100$). The only variable parameters that are accessible to the agents, are their weights, $w_{ij}$, which agents attribute to their neighbours. Similarly, as used earlier in the analysis of Glinton et al.’s model, for the stage of model exploration we assume that agents attribute the same common weight towards all of their neighbours, $w = w_{ij} \forall i \in A, j \in D_i$. We vary this common weight and explore behaviour of the metrics. Depending on the value
of the common weight, Figure 3.8A shows the number of agents that are expected to support the correct or incorrect opinions, while Figure 3.8B shows the metrics on the same scale of weights.

Since the designs of both models are reasonably close, their general behaviours are similar. As in Glinton et al.’s model earlier, we can observe the stable mode when the weights are too low to share opinions in the system, the unstable mode when early, and potentially inaccurate, opinions are shared on a large scale, and the critical mode when the accuracy is maximised. However, the models do not maximise their metrics with the same critical weights \( w_c \), and Glinton et al.’s model reports significantly higher accuracy improvement. The main difference in models lies in the fact that the sensing agents in our model introduce new opinions directly into the system without local filtering. This enables us to focus our study on the impact of collective behaviour on the accuracy of consensus regardless of the methods by which agents in a particular system acquire their observations. In contrast, the accuracy improvement in Glinton et al.’s model is a combination of effects of collective behaviour and local filtering, when sensing agents aggregate several observations before forming their opinion. Therefore, both metrics
Chapter 3 Modelling Collective Behaviour in Opinion Sharing

Figure 3.9: Sources of the accuracy improvement in our and Glinton et al.’s models. Local filtering in our model is absent, however, the sensing agents may report opinions formed under the influence of their neighbours, which result in a small difference in accuracy between observed and shared opinions (too small to be noticeable, shown as a bold error bar).

report higher results for Glinton et al.’s model, while results showed by our model give us answers to our research problem of exploiting properties of collective behaviour.

More specifically, despite the accuracy of the new opinions introduced into the models being the same, $r = 65\%$, the agents in Glinton et al.’s model fuse a number of new opinions locally, before forming their own opinions and sharing them with their neighbours. Therefore, the real accuracy of the opinions introduced into the system by the sensing agents is much higher. In our experiment we measure ‘Glinton’s input’ which is the accuracy of the opinions reported by sensing agents in Glinton et al.’s model. As can be seen in 3.8B, ‘Glinton’s input’ is much higher than the expected real accuracy of introduced observations $R_{\text{min}}^2$ due to the local filtering by the sensing agents.

As we expected, the results indicate that the reliability metric, $R_{\text{ratio}}$ is maximised when most of the agents do not form their own opinions at all. Specifically, it is maximised for our model when only $37\%$ of agents formed their own opinions, and $68\%$ of agents with opinions for Glinton et al.’s model. At the same time, our accuracy metric, $R$, is maximised when $97\text{-}99\%$ of agents form their own opinions, thus confirming our arguments above in favour of our accuracy metric.

In order to identify the sources of accuracy improvements in the models, we provide another experiment in which we compare the accuracy of the opinions reported by the sensing agents to their neighbours, $r_{\text{rep}}$, against the accuracy of introduced observations, $r$, and the accuracy of consensus, $R$. Specifically, we cover a broader range of settings and evaluate the models on 3 typical network topologies, 10 instances of each are generated with different random seeds. For each instance, we identify its critical weight $w_c$ and measure the highest accuracy of consensus, $R$, that can be achieved with it. Results presented in Figure 3.9 show that in our model the increment of the accuracy of consensus is only due to the collective behaviour which implements the distributed filter, $R - r_{\text{rep}} = R - r$. In contrast, the main source of accuracy improvement in Glinton et al.’s model
is the local filtering by the sensing agents, $r_{\text{rep}} - r$. Therefore, Glinton et al.’s model cannot be applied to our research problem. Specifically, in this model we cannot identify exactly how the collective behaviour influences the accuracy of consensus.

Having confirmed that our model has crucial differences to the existing models and that it is an appropriate choice for our research problem, next we analyse its properties in detail.

### 3.2 Computational Modelling Methodology

In our research we apply computational modelling (simulation) to analyse the properties of our opinion sharing model. The choice to use computational modelling is driven by several factors. Firstly, our model, being close to the models of social behaviour reviewed in Section 2.3, exhibits the same high level of complexity. Specifically, analytical tools that can be used to analyse dynamic processes on an arbitrary network topology have not yet been developed. The existing analytical approaches, such as the mean field theory, require full knowledge about the network topology generator and strong assumptions about the system properties, such as the scale of the system being infinitely large. Therefore, computational modelling for models of social behaviour was widely advocated with the development of powerful computers (Ball, 2012). Secondly, as we showed in the review of Glinton et al.’s model in Section 2.3.1, it is challenging to relate the performance metrics to the analysis of the processes of opinion dynamics. Specifically, correspondence of Glinton et al.’s analytical results to the maximisation of the reliability metric can be verified only by empirical evaluation. However, our empirical evaluation of their model also revealed the high influence of the model parameters on the position of the critical mode, when the metric is maximised. Therefore, we found that a careful selection of the model parameters is required in order for analytical results to coincide with empirical ones. Such an outcome highlights the discrepancy between the behaviour predicted by the analytical solution and the behaviour observed in more realistic settings for this type of model. However, following our research requirements, we aim to develop adaptive solutions that will operate in a wide range of settings. Therefore, we rely on computational modelling in our research.

Against this background, in this section we explain our goals in computational modelling. These are used to justify our choice of experimental setups, which we use to evaluate our model, and the decentralised algorithms designed for accuracy improvement which are presented in the following chapters. Specifically, we discuss in Section 3.2.1 how the research requirements can be evaluated, and in Section 3.2.2 we analyse the parameters of the model to select a number of the most indicative experimental setups. We conclude discussion on our methodology by describing the simulation process in Section 3.2.3.
3.2.1 Goals of the Computational Modelling

The ultimate objective of our research is to help agents in a large system find the correct opinion about the true state of the common subject of interest. The novelty of our problem lies in the imposed restrictions, where only a small number of highly uncertain sensing agents are present in the system. Crucially, as we identified in our motivating scenarios, due to their limited computational and communicational capabilities agents can often only share their opinions without any supporting information and quite often operate in sparse communication networks. To decompose our research problem, we identified in Chapter 1 a number of research requirements and developed the corresponding model in the previous sections. Following this list of research requirements, we now define the goals of our computational modelling experiments which are designed to explore the dynamic properties of the model. In particular, experiments that have to be conducted to test if it is feasible to achieve the following:

1. **Accuracy of Consensus**: We defined the accuracy of consensus as the expected correspondence of agents’ opinions to the correct opinion. We analysed the bounds on the accuracy of consensus, but simulation is required in order to investigate the level of accuracy which can be achieved in realistic settings. Moreover, our evaluation is expected to discover the factors that influence the model parameters on the position of the critical mode, when the highest accuracy is observed. Next, we need to test the hypotheses which were suggested by the previous research for Glinton et al.’s model. Specifically, the hypotheses that the value of the branching factor and the phase transition in the opinion dynamic indicate the position of the critical mode. Finally, we have to conclude which of these indicators are the most reliable to be exploited in designing our decentralised algorithms for tuning the system in the critical mode.

2. **Communication efficiency**: The main restriction on communication is imposed in the design of our model, which states that the agents are only able to exchange their opinions without any supporting information. Since we assumed that this limitation comes from the restricted capabilities of the agents, we have to investigate the communication expenses that are actually required by the system in order for agents to form their opinions. However, sharing opinions with the least number of messages in the system may result in slow convergence of the system to the consensus. Therefore, we additionally test the system performance with our convergence metrics.

3. **Adaptivity**: Most importantly, we need to investigate if accuracy can be improved in different parameter settings and how these settings influence the properties of the critical mode. In Section 2.2.2 we discussed the fact that the topological properties of a communication network have a significant impact on dynamic processes,
specifically in our case, on the opinion sharing processes. By conducting computational modelling, we have to analyse the specific influence of the network topologies on our model and, in designing decentralised algorithms for the improvement of accuracy in the following chapters, test their adaptivity. To test this requirement, we evaluate a number of chosen network topologies with variable parameters.

4. **Scalability**: Finally, we have to conduct an evaluation of systems of different sizes, starting with the smallest system which still exhibits the critical mode of behaviour and then steadily increasing the size until our results are no longer influenced by scale.

The outstanding requirement of ‘Robustness and Flexibility’ related to the decentralised algorithms for accuracy improvement, and we leave its analysis for the following chapters. Considering these goals of computational modelling, we now analyse the initial parameters of our model and select the experimental setups.

### 3.2.2 Experimental Setups

In this section we discuss the parameters of our model, their influence on the dynamic processes and our motivating for selecting their specific values. In particular, we differentiate the following components of the model and their corresponding parameters:

- **Agents**: A number of agents, \( N \), having their individual prior beliefs, \( p'_i \), and the common confidence bounds \( (1 - \sigma, \sigma) \) which when exceeded lead to opinion formation.

- **Sensing agents**: A small number \( N_s \) \( (N_s \ll N) \) of sensing agents are randomly distributed in the system, which share the same low probability of observing the true state of the subject of common interest, \( 50\% < r \ll 100\% \), and rate (periodicity) of introducing new opinions, \( \lambda \).

- **Communication Network**: Which defines a neighbourhood \( D_i \) of each agent and the dynamics of opinion sharing. The network can be characterised by its topological properties, such as the average path length, the clustering coefficient, the expected degree and its distribution. Since these properties are interdependent, we use the well recognised topology generators discussed in Section 2.2.2.3.

In the following subsections we briefly discuss the influence of these parameters and select their specific values for the evaluation process.
3.2.2.1 Agents

To examine how the properties of the model change with its scale, we conduct experiments for systems of \( N \in \{100 \ldots 10000\} \) agents. The upper and lower bounds were chosen empirically in order to show the smallest size of the system which exhibits the desired collective behaviour, and to investigate the upper size of system after which the behaviour does not change. Following our analysis of the accuracy metric, which showed that maximum accuracy increases with system size, we expect that higher accuracy can be achieved for larger systems.

The opinion formation of an agent \( i \) depends on its prior belief \( p'_i \), the confidence bounds of the decision rule, \((1−\sigma,\sigma)\), and the variable weights, \( w_i \), it attributes to its neighbours. Since we assumed that the confidence bounds are symmetrical, the specific choice of \( \sigma \) does not have a qualitative impact on the behaviour of the system. Only the agents’ weights have to be scaled accordingly, in order to repeat the same behaviour for the system with another value of \( \sigma \). Therefore, we assume that the confidence bounds are common for all agents in a system, and since the belief, \( p_i \), is defined on the range \([0,1]\), let \( \sigma = 0.8 \), which results in a wide range of agents’ beliefs \( p_i \in (1−\sigma,\sigma) = (0.2,0.8) \) corresponding to the ‘undetermined’ opinion following the definition of the agent’s decision rule (see Equation 3.4). This enables us to distribute the priors of the agents on a wide range of values, which introduces higher heterogeneity into the system.

The prior belief of the agent, \( p'_i \), is the parameter that encodes its preferences. Therefore, the distributions of the priors has a crucial impact on the system dynamics. The more diverse these priors are, the harder it is for the system to reach a consensus on which opinion is correct, and thus, the accuracy of consensus decreases. To illustrate this, we evaluate a number of systems of 1000 agents with different distribution of their priors. Specifically, we individually assign prior beliefs to the agents that are drawn from a normal distribution with different parameters: (A) the narrow distribution of \( p'_i \in \mathcal{N}(\mu = 0.5, s = 0.015) \); (B) the normal range of \( p'_i \in \mathcal{N}(\mu = 0.5, s = 0.09) \); and (C) the wide range of \( p'_i \in \mathcal{N}(\mu = 0.5, s = 0.5) \). In a case when a generated prior belief is out of the range of the confidence bounds, \( p'_i \not\in (1−\sigma,\sigma) \), a new value of the prior is chosen until it fits this range of the undetermined opinion.

The results showing the final opinions of the agents depending on a common weight for all agents, are presented in Figure 3.10. As expected, in case (C) the system cannot reach consensus. Crucially, there is no critical mode of collective behaviour when the system converges to the correct consensus opinion more often. Conversely, for the narrow distribution of priors (A) agents are less biased and we can observe the wider range of weights leading to the critical mode.

In order to examine the dependence of the critical mode from the system properties, and later to evaluate adaptivity of our decentralised algorithms, we choose more
challenging settings when the range of weights introducing the critical mode is very narrow. Therefore, we select for the empirical evaluation the distribution range (B), $p'_i \in \mathcal{N}(\mu = 0.5, s = 0.015)$.

Having defined all parameters of the agents, we now discuss the sensing agents which are responsible for introducing new opinions into the system.

### 3.2.2.2 Sensing Agents

New opinions are introduced into the system through a small number of sensing agents $N_s = 0.05 \cdot N$. However, new opinions have a low accuracy, $r = 65\%$, in which $r$ is the probability of observing the correct state of the common subject of interest. The small number of the sensing agents and their low accuracy was selected to reflect the statement of our research problem (see Section 1.3). Under such conditions the beliefs of the majority of the agents, $0.95 \cdot N$, are informed only by the opinions of their neighbours, because they cannot directly observe the state of the subject of common interest. At the same time, an opinion introduced by a single sensing agent is highly inaccurate, and thus, agents have to aggregate opinions from a number of sensing agents in order to form
the correct opinion. Thus, in this difficult setting we are able to focus more clearly on
the impact of collective behaviour on the accuracy of consensus.

We assume that the sensing agents are randomly distributed across the system. To
simulate a gradual introduction of new opinions, that corresponds to realistic settings,
every 10 opinion steps, defined as rate $\lambda = 10$, 10\% of the sensing agents are randomly
selected to make independent observations and introduce new opinions. This fixed rate
of the introduction of new opinions enables us to compare convergence of the model in
different settings in contrast to the dynamic rate in Glinton et al.’s model. The value
$\lambda = 10$ is selected empirically such that before initiating a new cascade, any previous
opinion cascade is likely to stop even on the largest sizes of communication networks.
Therefore the higher rate results in the same dynamics with correspondingly scaled time
axis. However, when $\lambda$ is smaller, a large opinion cascade initiated on the previous
round of observations may be supported or interrupted with newly introduced opinions.
Specifically, our empirical analysis showed that a lower rate of opinion introduction re-
sults in less stable behaviour in the system, when the range of the agents’ weights, which
result in the accuracy increase, becomes narrower. However, the qualitative properties
do not change. Therefore, we selected a rate such that we observe a more repetitive
behaviour. Crucially, since $\lambda > 0$, new opinions are introduced with realistic delays,
resulting in cascading behaviour in the model.

The opinions introduced by the sensing agents are shared between agents through the
communication network, the properties of which have a significant impact on the dy-
namics of the sharing processes.

3.2.2.3 Communication Network

We broadly discussed communication networks and their influence on the system dy-
namics in Section 2.2.1, and concluded that to simulate realistically complex topologies
we rely on the topology generators widely used in the literature:

- A random network, as a benchmark topology (see Section 2.2.2.3.1);
- A small-world ring network with a probability $p_{\text{rewire}} = 0.12$ of randomly rewired
  connections (Newman, 1999; also see Section 2.2.2.3.2);
- A scale-free network with a clustering factor $p_{\text{cluster}} = 0.7$ (Holme and Kim, 2002;
  also see Section 2.2.2.3.3);

We ensure that all our generated networks are connected into a single system. Specif-
ically, in this setting a single new opinion can be shared between all agents and can
lead to a consensus. Our early empirical study revealed that the dynamic processes of
opinion sharing on directed networks exhibit similar patterns of collective behaviour as
on undirected networks. Therefore, to simplify our earlier notation without losing the
generality of the results, we consider only undirected networks.

To evaluate the stability of collective behaviour, and later the adaptivity of our methods
of accuracy improvement, we consider a number of network instances with the expected
degree, \( \langle d \rangle \in \{6, 8, 12, 50, 100\} \). The values of the expected degree are chosen such
that our experiments cover sparse topologies, which are expected to be more sensitive
to the cascading behaviour, and to compare their performance with dense networks.
Specifically, in sparse networks, \( \langle d \rangle \ll N \), the agents are unlikely to have more than one
sensing neighbour, and the correct opinion has to be found on the system scale rather
than by each agent individually. The lower boundary of the expected degree, \( \langle d \rangle = 6 \),
is chosen such that our topology generators can generate a large connected topology in
a reasonable time. The upper boundary, \( \langle d \rangle = 100 \), is chosen such that for the smallest
size of a network, \( N = 100 \), we generate a fully connected network.

For each type of the network, which is defined by its topology generator, expected degree
and size of the system, we generate 10 instances with different random seeds. Thus, in
our study of the model performance in different settings, we avoid random biases by
analysing the averaged results. Since the topology of the communication network has
a significant influence on the dynamic processes, we summarise the properties of all
network instances we use in our empirical analysis in Figures 3.11 and 3.12. Specifically,
as we identified earlier in the literature review in Section 2.2.1, the average shortest path
length, \( \langle l \rangle \) (Equation 2.5), and the clustering coefficient, \( \langle C \rangle \) (Equation 2.9), are the
most indicative metrics on the sharing processes in a network.

In more detail, in our definition of the convergence metric we showed that the average
shortest path is the minimal number of steps required to share an opinion on the system
scale (Equation 3.14). Following this, the results presented in Figure 3.11 show that
the scale-free topology is expected to converge to consensus quicker than others. At the
same time, the small-world topology has the highest value of the average shortest path
length due to the properties of its generator, which rewires a ring network with a large
\( \langle l \rangle \) into a network with the small-world properties. Generally, all topologies exhibit an
increase in \( \langle l \rangle \) with the size of the network, which intuitively leads to an increase in the
convergence time for large systems.

The clustering coefficient, \( \langle C \rangle \), showed in Figure 3.12 indicates the connectivity between
the neighbours of an agent. If the clustering is high, agents form local groups in which
they are more likely to form the same opinion. This is due to the double counting
fallacy (see Section 2.3.1), when an opinion from the same sensing agent may arrive
multiple times via different routes, and an agent’s belief that its own opinion is correct
becomes much stronger than it should. Therefore, the value of the clustering coefficient
has a direct influence on the sizes of opinion cascades in the system, and speed of
its convergence to consensus. Since this is a crucial component of the process of
3.2.3 Simulation Process

To ensure that the results we observe are statistically significant, we simulate each set of parameters of the model over $|M| = 50$ independent opinion sharing rounds. After $|M|$ rounds we measure all our metrics and their standard errors.

opinion dynamics, we will investigate how the clustering is related to the values of our performance metrics.

Having discussed all initial parameters of the model, and their values that we are going to consider in the computational evaluation of our model, we now clarify the process of its simulation.
Table 3.1: Experimental setups for the model evaluation

<table>
<thead>
<tr>
<th>Model parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agents’ aggregation function</td>
<td>$f(\cdots)$</td>
<td>${\text{Bayesian, Weighted sum}}$</td>
</tr>
<tr>
<td>Number of agents</td>
<td>$N$</td>
<td>${100 \ldots 10000}$</td>
</tr>
<tr>
<td>Network topology</td>
<td>-</td>
<td>${\text{Random, Scale-free, Small-world}}$</td>
</tr>
<tr>
<td>Expected degree</td>
<td>$\langle d \rangle$</td>
<td>${6, 8, 12, 50, 100}$</td>
</tr>
<tr>
<td>Fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agents’ priors</td>
<td>$p'_i$</td>
<td>drawn from $\mathcal{N}(\mu = 0.5, s = 0.09)$</td>
</tr>
<tr>
<td>Agents’ confidence bounds</td>
<td>$(1 - \sigma, \sigma)$</td>
<td>$(0.2, 0.8)$</td>
</tr>
<tr>
<td>Number of sensing agents</td>
<td>$N_s$</td>
<td>0.05 · $N$</td>
</tr>
<tr>
<td>Accuracy of introduced opinions</td>
<td>$r$</td>
<td>65%</td>
</tr>
<tr>
<td>Rate of opinion introduction</td>
<td>$\lambda$</td>
<td>every 10 steps</td>
</tr>
<tr>
<td>Number of introduced opinions</td>
<td>$\Lambda$</td>
<td>$3 \cdot N_s$</td>
</tr>
<tr>
<td>Number of opinion sharing rounds</td>
<td>$</td>
<td>M</td>
</tr>
</tbody>
</table>

Moreover, in order to achieve unbiased results, we randomly choose the true state $b^m \in B$ of the common subject of interest before every opinion sharing round, $m$. Following this, each agent initialises its opinion $o_i^{k=0} = \text{undetermined}$ and belief $p_i^{k=0} = p'_i$. In order to simulate potential changes of opinions of the sensing agents, each opinion sharing round stops after introducing $\Lambda = 3 \cdot N_s$ new opinions into the system. We assume that this is the maximum number of observations that can be made about the state of the common subject of interest. Its value is chosen empirically such that even the largest systems converge to consensus and, at the same time, are not likely to change their opinions to new later-arrived opinions.

Table 3.1 summarises our choice of model parameters and thus, defines the experimental setups. Specifically, we identified 4 variable parameters in our empirical evaluation, the first of which, type of aggregation function, is defined in the model description. In order to avoid presenting detailed results for each set of parameters, in the next section we study the behaviour of the system when fixing the network topology to the scale-free generator, the size to $N = 1000$ agents with the expected degree to $\langle d \rangle = 8$. Following this we focus on the analysis of the critical mode in the full range of the experimental setups.

### 3.3 Social Dynamics in the Model

As discussed in Section 3.1.3, Glinton et al.’s model, exhibits an interesting collective behaviour in which the accuracy of consensus dramatically increases. Properties of this critical mode of the model dynamic rely on the fact that opinions are shared in the form of opinion cascades in which a single new observation may trigger a large number of agents to change their opinions, resulting in a sudden change in the system’s state (Bikhchandani et al., 1992). Specifically, in the critical mode, more frequent and
smaller cascades share opinions between limited numbers of agents. Furthermore, when the groups of agents sharing the same opinion coincide, less frequent but larger cascades occur and share this locally supported opinion on a global scale. Such behaviour implements a distributed opinion aggregation procedure, relying solely on the properties of the opinion sharing process. In such cases, the weights between the agents is the key parameter which regulates the sharing process and thus, impact the distribution of sizes of opinion cascades.

Unfortunately, it was shown that generally the critical weights, which introduce this critical mode, cannot be predicted (Glinton et al., 2009; Pryymak et al., 2012). Specifically, it was identified that the range of the critical weights is very narrow and highly dependent on the parameters of the system. Moreover, when a system has a complex topology of its communication network, this problem cannot be reduced to the averaged model in terms of mean field theory (Flyvbjerg et al., 1993) to allow its analysis (Glinton et al., 2010a). We have briefly pointed out the presence of the critical mode in our model and discussed its properties in Section 3.1.3, in which we compared our model with the existing ones. Now, we analyse these types of model dynamics in finer detail.

In particular, in this section we study behaviour of the model depending on the single control parameter which is accessible to the agents – the weight they attribute to the opinions of their network neighbours. Additionally, in order to identify the most influential factors in the accuracy of consensus and to indicate the settings with the highest accuracy, we compare two different agent designs (as discussed in Section 3.1.1).

It is computationally infeasible to evaluate the model performance with all the possible sets of agents’ weights, and so we make an assumption that all agents attribute the same common weight to all of its neighbours. This assumption was made above in the comparison of our model to the existing one and we showed that it was made before by Glinton et al. Our goal here is to identify modes of collective behaviour in our model, analyse when the accuracy of consensus is improved and identify which metrics indicate this mode.

In the following results we present an analysis of systems of $N = 1000$ agents on a scale-free topology with the expected degree $\langle d \rangle = 8$ (the rest of the parameters follow our selection in Table 3.1). We get broadly similar results within the full range of our experimental setup. Specific choice of the parameters mainly influences the position and the share of the critical mode which we discuss in detail in Section 3.4. Here, and in the following results all metrics are shown as averages with the error bars representing the standard error of the mean.

In our first experiment we analyse final opinions of the agents and the accuracy metric to identify the weights which induce the critical mode of behaviour.
3.3.1 Critical Mode with the Highest Accuracy of Consensus

In this series of experiments we examine the influence of the agent designs and the agents’ weights on the accuracy of consensus, which is the most important metric for the goal of our research. Specifically, we consider two different agent designs, based on the Bayesian aggregation rule and the weighted sum aggregation rule, which we introduced in Section 3.1. These two distinct aggregation rules were proposed in order to investigate the influence of agents on the collective behaviour of the system. Specifically, Figure 3.13A shows the final opinions for the systems operating with these two cases depending on the common weight they attribute to each other. Figure 3.13B presents our accuracy metric on the same scale. As can be seen, our accuracy metric directly follows the number of agents that have formed the correct opinion.

Crucially, these results show that the accuracy of consensus can be higher than $R_{\text{min2}}$, which is the accuracy of the system informed by a single sensing agent. This implies that the agents form their opinions relying on a number of sensing agents. Thus, in this critical mode the system exhibits a collective behaviour in which the agents are organised into a distributed opinion filter. As noted earlier, in this mode agents share new opinions in smaller groups and when two groups supporting the same opinion overlap, it is likely that a large opinion cascade will propagate it to the rest of the agents. To study these
dynamics, we later provide an additional analysis of the opinion cascades after examining our performance metrics. However, we can already confirm that the model exhibits the critical mode of collective behaviour in which the accuracy of consensus is significantly increased regardless of the agent design. This encouraging result enables us to approach our research problem in the following chapters.

Besides the critical mode, we observe two other modes of behaviour. When the system operates with the weights lower than critical, \( w \lt w_c \), the agents cannot form their own opinions and share them, since they do not form strong beliefs. This mode is known as the stable mode of the system. In this mode the accuracy follows our lowest bound \( R_{\text{min1}} \) (see Equation 3.10), since the agents do not share their opinions and only sensing agents form opinions with the expected accuracy of their observations, \( r \).

Conversely, the system is in the unstable mode when the weights exceed the critical ones, \( w \gt w_c \), and the accuracy of consensus is equal to the accuracy of an opinion introduced by a single sensing agent, \( R_{\text{min2}} \) (see Equation 3.11). In this mode the agents adopt the first opinion introduced into the system and share it on a large scale. By doing so, they aggregate the same opinion reported back from their neighbours and become sufficiently overconfident in their private beliefs leading to changing opinions when new observations arrive. Thus, we do not observe the distributed aggregation process in this mode, and despite the agents reaching a consensus, they do not benefit from the presence of a number of sensing agents.

The critical mode, when the accuracy of consensus is maximised, is the most interesting to analyse from the perspective of our research aims. In order to focus on its analysis, we define the critical mode as a range of model parameters which deliver at least 95% of the highest accuracy of consensus, which can be observed in an empirical evaluation of the system. In the results presented in Figure 3.13 and in the following figures of this section, we highlight the critical mode for both agent designs.

Most importantly, the critical mode is present for both agent designs. Despite the critical weights, \( w_c \), being different for our agent designs, the shapes of the accuracy plots repeat each other. This indicates that specific agent design influences specific values of the parameters which result in an accuracy improvement, but the properties of the collective behaviour do not depend on the specific aggregation function employed by the agents. Moreover, this intuitively expected result highlights that agent design should be considered in any analytical prediction of the model behaviour. However, the analytical prediction of the critical weights in Glinton et al.’s model (see Section 2.3.2) does not take into account the aggregation function employed by the agents. Thus, there is a clear need to improve the existing solution.

Finally, by analysing the critical state for both agent designs in Figure 3.13, we can see that the number of agents that hold the undetermined opinion at the end of a simulation is a promising indicator on the critical mode. Specifically, the number of agents that
do not form an opinion during the round significantly drops in a transition from the stable to the critical modes. In the critical mode itself, for the critical weight $w_c$, the expected share of agents with an undetermined opinion varies in a range of 1...5%. At the same time, in the unstable mode and in its transition phase, all agents form their opinion. Thus, if over a number of opinion sharing rounds a small share of agents (which can be different for each round) do not form their opinions, then the system is likely to operate in the critical mode. This is a global quantity and agents do not have access to it directly, however it leads to insights about the model behaviour and we test this hypothesis in the following sections. Before this, however, we analyse the rest of the metrics in the same settings, starting with the communication expenses.

### 3.3.1.1 Communication Expense

Since communication restrictions are imposed by our research problem, we investigate the communication expenses that are actually required by the system to operate in the critical mode. To this end, Figure 3.14 shows the number of messages which are transmitted in different modes, each carrying a single opinion between two neighbouring agents. Following our expectations, the communication expense metric runs in an opposite manner to the number of agents with the undetermined opinion in the system. Crucially, this number does not rise with the increase of the common weight, which suggests that agents are not likely to change their opinions many times and communication in the system is bounded. However, communication in the critical and unstable modes is slightly higher than the minimal communication, $U_{\text{min}}$, required to share an opinion in a single opinion cascade on the scale of the whole system. This indicates that $\frac{\max(U) - U_{\text{min}}}{N\langle d \rangle} = 6.2\%$ of the agents (where the expected degree, $\langle d \rangle$, is the number of messages that the agents communicate following an opinion change) do change their opinions in these modes in favour of the consensus, disregarding the early opinion they have adopted from a nearby sensing agent.
Unfortunately, sharing opinions with the number of messages approaching a minimal value may result in a slow convergence of the system to consensus. Therefore, next we analyse our convergence metrics.

### 3.3.1.2 Convergence to Consensus

Figure 3.15 presents the convergence metric. This metric is not defined in the stable mode, since the system does not reach a consensus, as we identified in the analysis of the accuracy metric in Figure 3.13. Following our hypothesis from the previous section, in the critical mode the system converges to the consensus much slower. Specifically, to reach a consensus in this mode, the system requires a two orders of magnitude higher number of steps than minimal, which is equal to the average shortest path: \( C_{\text{min}} = \langle l \rangle = 3.21 \). This time is required for all sensing agents to make their observations and report new opinions to the rest of the system.

Conversely, in the unstable mode, the convergence metric quickly approaches the theoretical minimum: \( C_{\text{unstable}} \approx 4.7 \). This confirms our earlier statement, that in this mode the early opinions are adopted by the system and that it does not benefit from the presence of a number of sensing agents.

Similarly as in the communication and the accuracy metrics, we do not observe a significant difference in the behaviour of the convergence metric for different agent designs. Again, this is a promising sign which indicates that the specific decision process of an agent does not influence the properties of the collective behaviour we examine.

### 3.3.2 Opinion Dynamics

In the existing analysis of accuracy improvement in an opinion sharing model similar to our own, Glinton et al. suggested that the branching factor, \( \alpha \), which is the expected number of neighbours that change their opinions following the change of an agent’s
Chapter 3 Modelling Collective Behaviour in Opinion Sharing

opinion, is a reliable indicator on the critical mode in their model (see Section 2.3.2). Specifically, their theoretical analysis suggested that in the critical mode $\alpha = 1$, and the distribution of the sizes of opinion cascades follows the power law. Considering that our and Glinton et al.’s models have crucial differences, which we identified in Section 3.1.3, we need to investigate if their results apply to our model.

To this end, Figure 3.16 shows the expected branching factor in our experimental setup, which indeed is close to 1 in the critical mode, despite being slightly lower when the accuracy is maximised: $\alpha_c = 0.84 \ldots 0.85$ ($0.65 \ldots 0.87$ for other network topologies). This suggests, after an agent changes its opinion in the critical mode, an expectation that one neighbour will adopt this opinion as well. In order to analyse if the same result is achieved in the full range of experimental setups, we additionally analyse the branching factor in the following section once more.

However, in so doing we discovered that the scale-invariant dynamics in opinion sharing, in which the size of opinion cascades are distributed by a power law, are not observed in the critical mode. Specifically, scale-invariant dynamics, when opinion cascades of all sizes are expected to occur, are expected to exhibit the highest variance. The results showing the variance of opinion cascade sizes are presented in Figure 3.17. The corresponding maximisation of the variance in opinion cascade sizes is observed in the unstable mode, which suggests that scale-invariant dynamics in opinion sharing are not observed in the critical mode in our model. Moreover, we investigated the moment when the variance is maximised and discovered that the power law cannot be fitted to the observed distribution of the opinion cascade sizes. Therefore, we cannot rely on the existing analysis of Glinton et al.’s model and have to develop new methods to find the model parameters which indicate the critical mode.

To conclude, the analysis of a sample experimental setup confirmed that our model exhibits the critical mode when accuracy is improved. Additionally, we showed that the specific decision rule employed by the agents does not have a qualitative influence on the properties of collective behaviour. Therefore, in the following section, which addresses
a wider range of experimental setups focusing on the properties of the critical mode, for the sake of brevity, we only analyse agents that adopt the Bayesian aggregation function.

3.4 Analysis of the Critical Mode

In this section we provide a wider analysis of the metrics by varying the rest of the parameters in our experimental setup. Specifically, we vary the network topologies, their expected degrees and the size of the system. The main goal of studying these empirical results is to investigate if accuracy can be improved in different, realistic settings, and how these settings influence the properties of the critical mode. Moreover, we intend to verify which metrics are reliable indicators of the critical mode, when the accuracy of consensus is improved.

Therefore, all results presented below correspond to the values of the metrics in the critical mode. Specifically, for each instance of the model we empirically find the common critical weight $w_c$ when the accuracy is maximised, by exploring a range of possible weights in a similar way to the previous section. Then, we take all measurements in the system with agents attributing $w_c$ to their neighbours.

Since the experimental setup covers a wide range of parameters, we present only the most interesting results in the following sections. However, as we identified in the previous section that the properties of different agent designs are very similar, in this section we focus only on analysis of the model with the Bayesian aggregation rule. We group metrics by the parameter which is the most influential on their behaviour. In the first set of experiments we analyse systems of $N = 1000$ agents with a variable network degree. Following this, we analyse the influence of systems scale on the critical mode by varying the number of agents, fixing the expected degree $\langle d \rangle = 8$. 

Figure 3.17: Variance in opinion cascade sizes depending on the common weight ($w_c$ and highlighted areas follow the description of the Figure 3.13)
Chapter 3 Modelling Collective Behaviour in Opinion Sharing

3.4.1 Accuracy Improvement

The first result, presented in Figure 3.18, confirms the critical mode when accuracy is significantly improved for all network topologies and their expected degrees. Specifically, we observe a significant improvement of the accuracy of consensus in comparison to the accuracy of a single sensing agent, denoted as $R_{\text{min2}}$. This result confirms that in the critical mode the system is organised into collective behaviour when inaccurate opinions are filtered out during the process of opinion sharing.

Notably, with increase of network density, which depends on the expected degree, the accuracy improves even further. This can be explained by analysing the network properties (see Figure 3.11) which suggest that the average shortest path decreases, and thus, the agents in the system become closer to the sources of new opinions, which are the sensing agents. This result confirms that network topology has a significant impact on the performance of our model. However, we also should note that the highest accuracy highly fluctuates for different network instances.

Since the weights, which the agents attribute to each other, is the only parameter that can be tuned, the most important question is the value of the critical weight, $w_c$.

3.4.2 Critical Agents’ Weights

In Figure 3.19 we show the value of the critical weight for systems with different network topologies (see Appendix A, Figure A.1 for additional results). These results indicate a clear dependence between the expected degree of the network and the value of the critical weight. With a higher number of neighbours (which corresponds to the expected degree)
an agent receives more opinions and thus, has to use a smaller weight to aggregate them in order to cross the confidence bounds. Subsequently, the agent is able to aggregate a higher number of opinions before making its own opinion and dense topologies with large expected degrees exhibit higher accuracy (see Figure 3.18).

Despite the results being very close, it is challenging to predict the value of the critical weight just by knowing the expected degree. After discussing our performance metrics, we later develop a benchmark to prove this claim for the full range of experimental setups.

### 3.4.3 Communication Expense

The next performance metric measures the number of messages that are communicated in the critical mode. Specifically, Figure 3.20 shows the strong dependence of the communication in the system on network density. In particular, agents have to communicate their opinions to a higher number of neighbours and communication expense correspondingly increases.

Crucially, in the critical mode we do not observe a significant deviation from the minimal communication, $U_{\text{min}}$ (Equation 3.12). This result indicates that most of the agents form their opinion only once and do not revise it. Thus, decisions on supporting the correct opinion are made in an intersection of small, local groups of agents which have already formed their opinions.

From these results we can see that communication might be even lower than the minimal, $U_{\text{min}}$, which is the number of messages required to share a single opinion to all agents. This implies that some of the agents do not form their opinions and thus, they do not
communicate. Moreover, the whole system might not converge to the consensus on some of the opinion sharing rounds. This, in turn, decreases the average result. Given this, we now analyse the convergence to prove this and show wider results.

### 3.4.4 Convergence to Consensus

Our evaluation reveals that the critical mode converges to consensus in 95% of the opinion sharing rounds. For the rounds when consensus has been formed, Figure 3.21 shows the convergence time, which is the expected time step when the system forms a consensus. This acts as an indicator on the timeliness of the agents’ opinions.

As can be seen, despite the significant variance of the individual experiments, the average value does not exhibit a clear dependence on the expected degree for the random and the small-world topologies. However, for the scale-free topology we observe that the convergence time increases for dense topologies. This observation cannot be explained only in terms of the network properties we analysed in Section 3.2.2.3, stressing the considerable influence of complex topologies on the performance of our model.

Along with the model performance metrics, we next analyse the branching factor which may indicate the critical mode.
3.4.5 Branching Factor as Indicator on the Critical Mode

In the previous section, we tested the hypotheses which were suggested by the previous research for Glinton et al.’s model. Specifically, that a value of 1 for the branching factor and the scale-invariant dynamic in the opinion sharing process indicate the critical mode. We concluded that the scale-invariant dynamic is not observed, while the value of the branching factor requires analysis in the wider experimental setups.

Given this, in Figure 3.22 we show the effective branching factor for different network topologies and network degrees (see Appendix A, Figure A.2 for additional results). High precision in the case with a random network might be a promising indicator here. However, our results for other topologies highlight that the branching factor is not an indicative measure of the critical mode for complex topologies. This experiment also explains a high sensitivity to the settings of the existing solution for finding critical weights in a distributed fashion, the DACOR algorithm, which we discussed earlier in Section 2.4.

Crucially, our results show that the branching factor is not a reliable indicator of the critical mode for our model and alternative indicators should be developed.

Next, we analyse the influence of system size on the model’s performance.

3.4.6 Influence of the System Size

For the last set of experiments we conduct an evaluation of systems of different sizes. In particular, the upper and lower bounds on the system size were chosen empirically in order to show the smallest size of the system which still exhibits the critical mode.
of behaviour, and to investigate the upper size of system after exceeding which the behaviour does not change. To provide an adequate comparison of the model’s performance, in these results we evaluate network topologies with two radically different expected degrees: $\langle d \rangle = \{8, 100\}$.

The results in Figure 3.23 show the accuracy of consensus depending on the system size. Following our analysis of the accuracy metric, we know that accuracy increases with system size, which is confirmed by the empirical results. Thus, the outcome of our choice of the bounds suggests that systems with $N = 100$ agents and less do not clearly exhibit the critical mode of behaviour by approaching the minimal bound, $R_{\text{min}2}$. Conversely, systems with more than $N = 5000$ agents do deliver a similar level of accuracy which suggests that further increases in system size will not lead to a change in accuracy. Crucially, in the cases of random and dense scale-free networks, we observe that the accuracy of consensus may reach the theoretical maximum of a centralised system, $R_{\text{max}}$. This indicates a high efficiency of the decentralised opinion aggregation implemented by the properties of collective behaviour in the critical mode.

Building on this, Figure 3.24 shows how system size influences the critical weights. Notably, systems with a smaller degree exhibit a higher sensitivity to the system parameters and thus, variation of the critical weight. At the same time, systems of dense networks are more predictable and their critical weights exhibit dependence, mainly on system size.

Finally, Figure 3.25 shows the convergence metric for different system sizes. Since all topologies exhibit an increase of the average shortest path length, $\langle l \rangle$, with the size of network, this intuitively leads to an increase of the convergence time for large systems. Still, this experiments confirms that in the critical mode the system converges to the
consensus slowly and the convergence time is at least two orders of magnitude higher than its minimum value $C_{\min} = \langle l \rangle = 2.8$.

### 3.5 Centralised Selection of the Critical Agents’ Weight

As we discussed in Section 3.4, we cannot analytically predict the critical weights which result in the highest accuracy of consensus. Therefore, we chose the computation modelling approach to explore the properties of the model empirically. Relying on our empirical exploration of the experimental setup, in this section we design a number of benchmark methods for improving the accuracy of consensus. In order to make the empirical exploration of the parameters feasible, we assume that all agents use the same common weight, $w_{ij} = w \ \forall i \in A$, which is defined by a centralised authority. These
methods will then be the benchmarks for the decentralised solutions for accuracy improvement, which we offer in the Chapters 4 and 5.

In more detail, we offer three benchmarks for improving the accuracy of consensus. Specifically, we consider the following scenarios:

- When the information about the system is perfect. Thus, we can simulate the system off-line and choose the best weights in a centralised manner.

- When we know all parameters of the system. Relying on this we can predict the most beneficial weights by analysing systems with the same set of parameters.

- When we do not know the parameters. For this worst case scenario, we design a strategy of choosing agents’ weights which minimises the accuracy loss in comparison to a random guess.

In the following sections we describe these benchmarks and conclude with analysis of the accuracy of consensus which they achieve.

### 3.5.1 When Information about the System is Perfect

The first benchmark assumes that we have perfect information about the system parameters, which includes the exact topology of the communication network. We offer this benchmark in order to demonstrate:

- An off-line solution for improving the accuracy of consensus by empirically evaluating the system’s performance with a large number of different agents’ weights;
• The complexity of the problem and, thus, the need to develop a decentralised runtime solution;

Specifically, we can offer a simple approach to pre-tuning a system by empirically evaluating it with a number of weights and selecting the critical weight that delivers the highest accuracy $R$.

However, as we discussed in our computational modelling approach in Section 3.2, we assume that agents attribute the same common weight to each other. Specifically, it is computationally infeasible to evaluate all cases with agents attributing individually selected weights, since the number of such experiments is combinatorial in the number of agents and weights we should consider in the evaluation.

Therefore, to pre-tune a system as we did in Section 3.4, we need to perform a resource intensive empirical exploration of the system performance with a common weight $w_i = w \forall i \in A$, where $w \in (0.5, 1)$, with a sample step of 0.05. Then we choose the weight $w_c$ at which the system exhibits the highest accuracy $R$. Clearly, this approach requires significant computational resources, since the system has to be evaluated over a number of sharing rounds, $|M| = 50$, for each possible value of weight in order to find the critical weight. Therefore, this empirical exploration cannot be performed at runtime and used in realistic settings, thus, alternative solutions are required.

### 3.5.2 When the System Parameters are Known

In most cases, it is unlikely that we can observe the exact topology of a large system. Therefore, in developing this benchmark, we consider a case in which all parameters are known, including the properties of the communication network, however, the exact topology is unknown and the individual tuning of a system is not possible.

In this case, we may evaluate a number of systems with the same parameters and the configuration of a network topology generator. Then, we generate a number of topologies and individually pre-tune each system as we described in the previous section. However, $w_c$ will vary significantly between different network instances, since the area of the critical weights is very narrow and sensitive to the system parameters. Therefore, in order to show that it is hard to predict the critical weight which delivers the highest accuracy, we provide a benchmark in which a system operates with the average critical weight, $\langle w_c \rangle$. This average critical weight is calculated by averaging individually selected critical weights for all systems, which were generated with the same parameters (including the parameters that are variable in our experimental setup, such as an agent’s aggregation function, the system size, the network topology and its expected degree).
This benchmark is designed to stress the high sensitivity of critical weight to the specific topology of a system. It would confirm the need for selecting the critical weight for each system individually.

3.5.3 When the System Parameters are Unknown

Finally, we introduce an additional benchmark which can be applied in a case in which we do not know the parameters of the system. In this case it is reasonable to select a common weight such that the system would be operating in the unstable mode. Thus, this approach is expected to deliver accuracy at the minimal level $R_{\text{min}2}$, which is significantly higher than the accuracy of the system in stable mode, $R_{\text{min}1}$. This benchmark indicates the improvement of accuracy that can be achieved by other methods in comparison to the unstable mode with the guaranteed level of accuracy.

3.5.4 Empirical Evaluation

The accuracy of consensus delivered by each of the benchmarks in our experimental setup is shown in Figure 3.26 (for fixed $\langle d \rangle = 8$). Note that our assumption of choosing a common weight for all agents, instead of selecting the weights individually, suggests that the benchmarks do not reach the optimal configuration. However, as our evaluation shows, systems relying on this approximation may exhibit a high accuracy of consensus with $R = 90 - 97\%$ for large systems with the random topology, approaching the theoretical maximum, $R_{\text{max}}$.

On the other side, the accuracy in the unstable mode is close to the analytically predicted $R_{\text{min}2}$. This bound is the accuracy of the system forming the consensus by adopting the first and thus, inaccurate opinion introduced into the system. More importantly, the discrepancy in the performance of systems with the average critical weight and the individually pre-tuned weight, confirms our statement that specific topology has a significant influence on the critical weight. Therefore, methods for individual selection of the critical weights should be developed. Finally, our benchmarks confirm again that the type of network topology influences the performance of a system.

3.6 Summary

In this chapter we presented our opinion sharing model and conducted its analysis. Our results showed that the model exhibits the mode of collective behaviour in which the accuracy of consensus is significantly increased. This encouraging result enables us to approach our research problem of improving the accuracy of consensus in large decentralised systems with limited communication. Specifically, we intend to solve this
problem by exploiting the properties of the identified critical mode. Its analysis suggested the metrics that indicate whether a system operates in the critical mode. Relying on this, in the following chapters we will develop decentralised algorithms for accuracy improvement in large multi-agent systems.

In more detail, in this chapter we designed a novel opinion sharing model, which is the first to quantify the exact impact of collective behaviour on the accuracy of consensus. We showed the differences of our model from the existing ones and confirmed that we successfully addressed the shortcomings in the model closest to ours, offered by Glinton et al.

Next, we offered metrics to measure compliance with our research requirements and provided analytical bounds on their values. Following this, we explained our motivation for adopting a computational modelling approach. After examining the model’s parameters we selected a representative range of experimental setups. Our analysis of our model confirmed the presence of the critical mode when the accuracy of consensus is improved,

Figure 3.26: Accuracy of consensus achieved by different benchmarks depending on the system size and topology ($\langle d \rangle = 8$, Bayesian aggregation function)
and crucially, suggested indicators of this state. In the wider analysis, which included
the full range of experimental setups, we examined variations of the parameters in the
critical mode. Crucially, we showed that the branching factor is not a reliable indicator
of the critical mode, which was advocated in the existing research. Moreover, we showed
that the phase transition in sizes of opinion cascades occurs in the unstable mode. This
suggests that in our model the scale-invariant dynamic in opinion sharing (the critical
phenomena) is not observed in the critical mode. Thus, we cannot build on the existing
analysis of Glinton et al.’s model and we have to develop new methods for finding the
model parameters which lead to the critical mode.
Chapter 4

Accurate Consensus with Anonymous Peers

Relying on the analysis of our opinion sharing model, we now solve our research problem. Specifically, we improve the accuracy of consensus in large decentralised systems with restricted communication. In order to do so, we offer a novel agent behavioural algorithm which exploits the discovered properties of the collective behaviour in our model. This is the first solution to meet our research requirements in the more difficult case in which peers are anonymous (Requirement 2a, Chapter 1). In this case the agents in a system cannot identify their peers, which are their main source of observed opinions, and thus, have to treat them all equally.

More specifically, we develop a novel decentralised algorithm, Adaptive Autonomous Tuning (AAT), which significantly improves the accuracy of consensus in comparison to the accuracy of opinions introduced into a system. The algorithm achieves a promising level of performance in large multi-agent systems with complex communication networks. It does so by independently helping each agent to weight the received opinions, such that the whole system self-organises into the critical mode of behaviour we identified in Section 3.3. In this mode a multi-agent system filters early and possibly inaccurate opinions by sharing them amongst small groups of neighbouring agents, which prevents overreaction. When several groups with the same opinion overlap, this locally supported opinion is shared on a large scale leading to a system-wide consensus. This opinion sharing pattern implements a decentralised aggregation of opinions from a number of different sources on the scale of a large system. Such an approach based on a specific mode of the collective behaviour overcomes the limitation of a single agent which cannot form an accurate opinion given its highly restricted view.

Crucially, AAT is the first solution that meets our minimal communication requirement. It operates successfully when communication is limited to sharing opinions without any supporting information. In contrast, the current state-of-the-art solution, DACOR,
communicates up to 4-7 times more service messages than is required to share opinions. We empirically evaluate AAT and show that it significantly outperforms DACOR, and approaches the highest centralised benchmark we introduced in the previous chapter. Specifically, using AAT, the accuracy of consensus reaches 75-93\% (probability of the correct consensus) given only 5\% of sensing agents that can make a noisy observations (only 65\% of which correspond to the correct opinion). This figure is significantly higher than the 65-75\% achieved by DACOR. At the same time, the performance of AAT is close to the 80-97\% for a system pre-tuned for the highest accuracy by an intensive empirical exploration of its parameters.

Moreover, AAT has lower operational costs and requires up to $5 \cdot 10^4$ times less agent actions, such as a message transmission or a weight change, than DACOR to achieve the beneficial self-organised mode. Additionally, we look into optimising its computational cost by offering a number of heuristics to replace computationally intensive stages. By doing so, we significantly reduce the algorithm’s search space and speed up selection of the best solution. The runtime of the improved AAT becomes 3-4 times lower than regular AAT. This new figure falls into the range of measurement error of the DACOR runtime. Crucially, this figure is much closer to the runtime of a static system, which does not employ a behavioural algorithm.

Finally, we show that AAT is the first decentralised solution designed to improve the accuracy of consensus in heterogeneous systems, which include faulty or indifferent agents that do not participate in the optimisation process. Specifically, AAT significantly improves the accuracy when up to 80-90\% of the agents in the system use fixed randomised weights instead of running the AAT algorithm. This implies that AAT is tolerant to this type of fault. Thus, it can potentially be used in existing large systems where it is impossible to update the behaviour of all agents simultaneously.

The remainder of this chapter is organised as follows. In Section 4.1 we present the core of our AAT algorithm. Following this, in Section 4.2 we look into improving its efficiency by bounding its search space and developing a heuristic approach that dramatically reduces its computational expenses. Then, Section 4.3 examines parameters of the algorithm and suggests the best choice for our experimental setup. With these parameters we evaluate AAT against the state-of-the-art DACOR algorithm, and the benchmarks offered earlier in Section 4.4. Finally, we conclude in Section 4.5 by discussing how the algorithm offered meets the research requirements.

4.1 The Autonomous Adaptive Tuning Algorithm

In this section, we present our Autonomous Adaptive Tuning (AAT) behavioural algorithm for improving the accuracy of consensus, $R$, defined in Equation 3.6. Specifically, AAT is designed to operate in large decentralised systems by exploiting properties of
their collective behaviour. In contrast to the DACOR algorithm (discussed in Section 2.4), our solution does not introduce communication overhead and communication is strictly limited to opinion sharing. Specifically, an agent running the DACOR algorithm communicates service messages to all its neighbours after it observed that any of them has changed its opinion. Therefore, following each opinion change, DACOR agents communicate up to $\langle d \rangle^2$ additional service messages, where $\langle d \rangle$ is the expected number of neighbours.

We address this shortcoming by developing a new algorithm that updates agents’ weights autonomously, relying on their local observations only. Moreover, our analysis of the model indicates that the techniques used in designing DACOR are not reliable indicators of the settings with improved accuracy (the branching factor in the critical mode of system dynamics is not equal to 1 as we analysed in Section 3.4.5).

In contrast, AAT is built on the observation that the accuracy significantly increases when the dynamics of opinion sharing is in the critical mode. The narrow range of weights that introduced this mode lie between the stable mode (when opinions are not shared) and an unstable one (when the first introduced opinion is propagated on a large scale). The critical mode creates a condition where the system does not overreact to incorrect opinions and the agents share opinions in smaller groups before a large cascade occurs. To reach this area of optimised parameters, AAT gradually tunes the weights of each agent individually.

The three stages of AAT, illustrated in Figure 4.1, are described in detail in the following sections. First, each agent running AAT populates a set of candidate weights to reduce its search space. This step is described in Section 4.1.2. Then, the agent estimates the fitness of the candidate weights after each opinion sharing round, as described in Section 4.1.3. Finally, Section 4.1.4 discusses how the agent selects a weight to use in the following round, considering how close its fitness is to the target value.

The most important question is the choice of the weights’ fitness function. As we identified in the model evaluation (Section 3.4), the model parameters have a significant influence on the position of the narrow range of the critical weights, which correspond to the critical mode of dynamics. Therefore, it is a very challenging task to find these weights in a decentralised fashion. Our algorithm attempts to satisfy this objective and in Section 4.1.1 we discuss its crucial component, which is an indicator of the critical mode.

### 4.1.1 Awareness Rate as an Indicator of the Critical Mode

In the previous chapter we showed that the scale-invariant distribution of the sizes of opinion cascades is not observed in the critical mode of our model (Section 3.3.2). Also,
we examined the branching factor and showed that it cannot be used as a reliable indicator of the critical mode (see Section 3.4.5). Thus, the indicators offered by the previous research, along with the DACOR algorithm, cannot be used to solve our problem.

To address this issue, in Section 3.4 we analysed which metrics indicate the critical mode regardless of the initial parameters of the model. We identified that in the critical mode the share of agents that form their opinions approaches 100%. To illustrate this, we plot in Figure 4.2 the expected share of agents holding the correct and incorrect opinions in different modes of model behaviour depending on the common weight agents apply to each other. Additionally, we plot the average awareness rate of the system, which is the probability of agents forming their opinions, which we denote as $\langle h \rangle$. This metric is one minus the share of agents holding an undetermined opinion at the end of a round:

$$
\langle h \rangle = \left(1 - \frac{|\{i \in A : o_i^{m} \neq \text{undetermined}\}|}{N}\right) \cdot 100\% 
$$

(4.1)

It is notable that in our results, the accuracy of consensus, which follows the share of agents holding the correct opinion, dramatically increases when agents use the minimal weight that enables them to form their opinions. When this condition is met, the agents form their opinions with $\langle h \rangle = 96\%$, which is slightly lower than the maximum. This
transition to the maximum value of the awareness rate indicates the change from the critical mode to the unstable mode of model dynamics. A similar pattern is observed for the model with other initial parameters such as its size, network topology or the decision rule employed by the agents.

Relying on this observation, we now offer a myopic indicator of the critical mode that can be calculated by each agent individually. Specifically, from its own perspective, a single agent \( i \) cannot determine when it has formed the correct opinion which corresponds to the correct state \( b^m \). However, it is important to know how often the agent forms its opinion. To measure this, we define an agent’s awareness rate, \( h_i \), as the proportion of opinion sharing rounds where the agent \( i \) held an opinion, rather than being undetermined, compared to the total number of rounds, \( |M| \):

\[
h_i = \frac{|\{m \in M : a^m_i \neq \text{undetermined}\}|}{|M|}
\]  

(4.2)

This metric can be calculated by each agent locally and we use it as the basis of our algorithm.

In more detail, the intuition behind our approach is that in order to form an accurate opinion, the agent has to gather as many of its neighbours’ opinions as possible before forming its own opinion. To do so, it has to use the minimal weight that enables it to form an opinion when all its neighbours have reported theirs. However, if all agents use the minimal weight and wait until all their neighbours form opinions, a deadlock results in which the opinion sharing stops. Therefore, each agent must apply a minimal weight to the received opinions which guarantees that the agent actually forms its own opinion and shares it further.
In terms of the model we can formalise this, such that in order to maximise the accuracy, \( R \) each agent has to:

- Form its opinion and thus, reach a high awareness rate (\( h_i \), the proportion of the rounds where the agent held an opinion rather than being undetermined) since the agents with undetermined opinions decrease the accuracy;

- Form the correct opinion given its local view. Following the intuition above, in order to do so, the agent has to form an opinion as late as it is possible to gather the maximum number of neighbours’ opinions.

To meet these conditions, the agent has to use the minimal weight that always leads to an opinion formation (\( h_i = 1 \)).

However, since the sensing agents introduce observations randomly, the opinion sharing dynamic in the critical mode exhibits stochastic behaviour. As a result, during some rounds opinions are not shared on a large scale and the agents’ awareness rates suffer. Therefore, to improve the overall accuracy and to find the exact position of the critical mode, each agent \( i \) has to compromise its own awareness rate, \( h_i \). Specifically, the agent has to find the minimal weight, \( w^i \) out of candidates \( W_i \) that delivers the target awareness rate, \( h_{\text{trg}} \), that is slightly lower than the maximum, 1. Formally, each agent solves the following optimisation problem:

\[
w_i = \arg \min_{w^i \in W_i} |h_i(w^i) - h_{\text{trg}}| \quad (4.3)
\]

where \( h_i(w^i) \) is the awareness rate that the agent achieves using weight \( w^i \). We analyse the impact of the specific value of \( h_{\text{trg}} \) on accuracy in the empirical evaluation of AAT in Section 4.3.1.

Having identified the indicator of the critical mode and the optimisation function that AAT solves, we now present the stages of our algorithm as illustrated in Figure 4.1.

### 4.1.2 Candidate Weights

Each agent \( i \) running AAT must initialise itself with a set of the candidate weights, \( W_i \), which reduces the continuous problem of selecting its weight, \( w_i \), from the range \([0.5, 1]\), to a discrete problem. In the optimal case, this set of candidates contains weights that correspond to the distinct dynamics of an agent’s opinion, since only the moments of opinion formation can be observed by its neighbours, and thus, influence the system. Later, in the optimisation of AAT in Section 4.2.1 we analyse this idea in detail.

However, reducing the search space to a small set of predefined candidates can be very challenging due to a number of reasons. For example, this problem is undefined when
an agent cannot identify a number of its neighbours (to analyse their possible opinion
dynamics) or when the underlying communication network is dynamic. In the latter
case, the set of optimal candidates may also vary.

Therefore, in the general case when conditions are unknown, $W_i$ should be populated
with weights drawn from the range $[0.5, 1]$ with a given step size, for example 0.01. This set of candidates might be larger than the optimal since it might contain redun-
dant weights or, conversely, it might be missing weights that encode some crucial cases
of an agent’s dynamic. Due to this, such an approach significantly increases the re-
quired computational resources and, as we shall see later, may slow the convergence of
the algorithm to the critical mode. Therefore, in the following section we address this
problem by only populating the set of candidates with weights using which an agent
would exhibit distinct opinion formation dynamics. In order to do so, we have to intro-
duce the assumptions that an agent knows the number of its neighbours and that the
communication network is static.

Now, having initialised the set of candidate weights, the agent has to select a weight to
attribute to the opinions of its neighbours. However, in order to do so, it has to estimate
how likely it is that each of the candidates will lead to the critical mode.

### 4.1.3 Dynamic Estimation of the Awareness Rates

In this section we discuss how the criteria the AAT algorithm uses to select a weight
from the candidates is calculated. As mentioned earlier, AAT is based on our observation
that the accuracy of consensus, $R$, is maximised when the agents attribute the minimal
weights to their peers which still enable them to share opinions on the system scale.
Formally, this involves solving the optimisation problem defined in Equation 4.3.

In order to solve this optimisation problem, the agent needs to calculate all awareness
rates, $h(w_i)$, that would be achieved by using each candidate $w_i \in W_i$. However, the
agents’ opinions are highly interdependent and the choices of an individual agent eventu-
ally affect the dynamic of the whole system. Therefore, awareness rates can only be
estimated empirically through a number of opinion sharing rounds.

According to the definition of the awareness rate, $h_i$ (Equation 4.2), it can be measured
only for the weight, $w_i$, that the agent currently uses. Thus, in order to update the
awareness rates of all the candidates, the agent has to record the sequence of the opinions
it has received. Then, by locally emulating this recorded opinion formation process for
each candidate weight, the agent is able to identify if it could have formed its own
opinion with a given candidate weight.

This approach is likely to require significant computational resources. Since our research
aim is to develop a computationally efficient solution, in the following optimisation of
AAT in Section 4.2.2, we develop heuristic criteria which do not require us to recalculate the opinion formation process. However, in order to do so, we have to assume that the exact moments of the agent’s opinion formation do not influence the dynamics of its neighbourhood, which is unlikely due to the high interdependency between agents’ opinions. Therefore, the most reliable approach to measure the awareness rate of a candidate is its direct evaluation over a number of opinion sharing rounds. Therefore, the strategy of selecting a candidate’s weight has significant impact on the convergence of AAT to the solution of its optimisation problem. Now, we discuss such strategies.

4.1.4 Weight Selection Strategies

The problem of selecting the best weight out of the candidates according to their estimated awareness rates resembles a standard multiarmed bandit (MAB) problem (Katehakis and Veinott, 1987). In the MAB problem, there is a machine with \(|w_i| \in W_i|\) arms (the number of the candidate weights in our case), each of which delivers a reward, \(h(w_i)\) (the awareness rate), that is independently drawn from an unknown distribution, when the machine’s arm is pulled. Given this, we can apply the following widely recognised MAB strategies (Vermorel and Mohri, 2005) to select the weight out of the candidates:

- **Greedy**: A benchmark that selects the weight, which has the awareness rate closest to \(h_{trg}\).
- **\(\epsilon\)-greedy**: Selects the weight closest to the target awareness rate with probability \(\epsilon - 1\), otherwise it selects a random one (let the random factor be \(\epsilon = 0.1\)).
- **\(\epsilon\)-N-greedy**: The same as above but the exploration factor decays over time as \((\epsilon - 1)/f(m)^2\) where \(f(m)\) is selected such that the random factor becomes insignificant after \(m > 150\) opinion sharing rounds.
- **Soft-max**: Chooses each weight with probability \(\frac{\exp(q(w_i)/\tau)}{\sum_{v=1}^{W_i} \exp(q(w_v)/\tau)}\), where \(q(w_i)\) is the distance between \(h_i(w_i)\) and \(h_{trg}\), and \(\tau\) is the damping factor that decays to 0 after \(m > 150\) of opinion sharing rounds.

The latter two strategies gradually decay their exploration over time. Following our note regarding the high interdependence of agents’ opinions earlier, a weight chosen by a single agent influences opinion dynamics in the whole system. Therefore, we expect the strategies with less dramatic changes in agents’ dynamics to converge to the solution in a smaller number of opinion sharing rounds and not to fluctuate around the solution. The exploration phase cannot be avoided completely since, as we discussed in the previous subsection, the awareness rate can be accurately measured only for the weight that the agent currently uses.
Algorithm 2 AAT: Hill-climbing strategy to select a weight

Function \text{ChooseWeight}(i, \epsilon = 0.05)

1: \( W_i := \langle \text{SortAsc}(W_i) \rangle \) \{in order to use position indexes\}
2: \( l := \text{getIndex}(w_i, W_i) \)
3: \text{if } l < |W_i| \text{ and } \hat{h}^m(w_i^l) < h_{\text{trg}} \text{ then}
4: \quad l := l + 1 \{ \text{increase the weight to the nearest higher candidate} \}
5: \text{else if } l > 1 \text{ and } \hat{h}^m(w_i^{l-1}) > h_{\text{trg}} + \epsilon \text{ then}
6: \quad l := l - 1 \{ \text{decrease the weight to the nearest lower candidate} \}
7: \text{end if}
8: \text{return } w_i^l

MAB strategies assume that the distribution of awareness rates is unknown, however its shape can be estimated. For the candidate weights, \( W_i \), sorted in ascending order the smallest weight, \( w_1^i \), requires more sequential updates to cross one of the confidence bounds, while the largest \( w_{|W_i|}^i \) requires less, and thus we expect \( h(w_1^i) \ll h(w_{|W_i|}^i) \). Consequently, awareness rates are distributed as a hill with a peak for the largest weight. Therefore, we offer an additional strategy that makes use of this observation:

- **Hill-climbing**: Select a weight to use in the next round from the closest candidate weights to the one currently used. Specifically, if the awareness rate delivered by the currently used weight, \( w_i \), is lower than the target \( h_{\text{trg}} \), the agent must increase the weight to the closest larger candidate. Conversely, the agent decreases the weight, if the closest lower candidate weight is estimated to deliver an awareness rate higher than the target.

Algorithm 2 presents the formal definition of the hill-climbing strategy. We introduce an additional hysteresis parameter, \( \epsilon \), in order to reduce the number of changes of weights even further. We expect this strategy to deliver the highest accuracy, since it introduces less change to the system dynamics during the exploration phase and therefore the awareness rates may be estimated with a higher accuracy.

To confirm this hypothesis and to show that AAT meets our research requirements, we provide an extensive empirical evaluation in the Section 4.3. However, as we noted in the algorithm description, some stages of AAT can be improved before the evaluation.

### 4.2 Reducing Computational Cost of AAT Algorithm

In the description of the first stage of our algorithm, which is the selection of the candidate weights, we offer a generic solution of populating the set of candidates. Specifically, the candidate weights are uniformly drawn from the allowed range. We noted that this solution is not optimal and it produces an excessive number of candidates. Most of these candidate weights result in the same behaviour of an agent, thus they are redundant.
This results in a higher number of opinion rounds required by AAT, since it has to estimate the awareness rates of a large number of the candidates. In turn, the system running AAT self-organises into the critical mode slower.

Nevertheless, this generic approach can be successfully applied in many challenging settings, such as anonymous our dynamic networks. However, if some additional information is available to an agent, such as the number of its neighbours, it can form a much smaller set of the candidate weights. Specifically, in Section 4.2.1 we illuminate the redundant candidates by offering a new approach to generate the optimal set of the candidate weights.

Following this, in Section 4.2.2, we address a weakness outlined in the second stage of AAT, which is responsible for estimating the awareness rates for all candidate weights. Specifically, we rely on the properties of agents’ decision rules and offer indicators to estimate the awareness rates of the candidates which were not used during an opinion sharing round. This helps us avoid the computationally expensive simulation of each individual candidate which we offered earlier.

### 4.2.1 Limiting the Set of Candidate Weights

In this section, we analyse the dynamics of an agent’s belief, \( p_i \), over the number of belief update steps \( k \) in order to limit the search space for each agent from the continuous interval \( w_i \in [0.5, 1] \) to a finite set of candidate weights. Specifically, we identify only those weights using which the agent exhibits distinct dynamics. In so doing, we rely on the assumptions that the number of neighbours is known to an agent and that the network is static. Otherwise, the more generic solution offered earlier in Section 4.1.2 should be used.

Consider that each agent, \( i \in A \), sequentially receives opinions from its neighbours and that these opinions may or may not be conflicting. For example, Figure 4.3 illustrates the sample dynamics of the agent’s belief, \( p_i^k \), curated by the Bayesian aggregation function. Here the agent initially receives a number of opinions from its neighbours that indicate that the correct state is ‘blue’ and meanwhile the agent forms the corresponding opinion for itself. However, the opinions that arrive later support the opposite state, ‘orange’, and after a number of updates the agent switches its opinion to support this the new opinion that the correct state is ‘orange’. Such dynamics indicate that the agent participated at least in two opinion cascades that were propagating conflicting opinions.

Considering the dynamics described above, note that during each step, \( k \), of its belief update, the agent has a number of opinions, \( u_i^k \), received from its neighbours that support an opinion \( b^{\text{op}} = \text{orange} \), and a number of received opinions, \( w_i^k \), that support the conflicting opinion \( b^{\text{op}} = \text{blue} \). Following this, during the whole opinion sharing round,
Chapter 4 Accurate Consensus with Anonymous Peers

Figure 4.3: Sample dynamics of the agent’s belief with marked steps when the agent changed its opinion. Starting from its prior, \( p_i' \), the agent updates its belief with 4 neighbours’ opinions that support ‘blue’ after which the agent sequentially receives 11 opinions supporting ‘orange’. As a result, the strongest support in this example is \( u^m_i = |4 - 11| = 7 \).

There are some belief update steps, \( k \), when an agent observes the strongest support in favour of one of the conflicting opinions, which is when the largest number of received opinions simultaneously support one of the conflicting opinions. We denote ongoing support as the difference between the received conflicting opinions, \( u^k_i - \pi^k_i \), and the strongest observed support during round \( m \) as:

\[
  u^m_i = \max_k |u^k_i - \pi^k_i| 
\]  

(4.4)

For example, in Figure 4.3, the strongest support is \( u^m_i = |4 - 11| = 7 \) and this is observed in the last belief update step.

When the agent observes the strongest support its belief is maximised or minimised and thus, the agent is most confident in forming its most accurate opinion given its local view. In order to form the opinion exactly when the strongest support is observed, the agent’s private belief, \( p^k_i \), has to match one of the confidence bounds, \( p^k_i \in \{1 - \sigma, \sigma\} \), (in our example in Figure 4.3 this implies that the agent’s belief should reach one of the confidence bounds and stay in the range \([1 - \sigma, \sigma]\)). Since the agent’s weight, \( w_i \), influences the dynamic of its belief, we can select two optimal weights that meet the described condition given a specific value of the strongest support, \( u^m_i \). If the agent’s weight is higher than optimal, the agent forms a less accurate opinion earlier than the strongest support is observed and becomes overconfident. Conversely, if the agent’s weight is lower than optimal, the agent is not able to form its own opinion given the observed strongest support.

Now, we discuss the possible states of the strongest support that the agent can observe. Then we develop a method to find the optimal weights for each case, that together form a set of the candidate weights in order to reduce the search space in the following stages of our algorithm.

The number of received opinions that support one of the conflicting beliefs, \( u^k_i \) and \( \pi^k_i \), is limited to the total number of the neighbours, \(|D_i|\). Following the definition of the strongest support in Equation 4.4, we can conclude that it is also limited by the number...
of agents’ neighbours:

\[ u_i^m \leq |D_i|, \quad u_i^m \in \{1 \ldots |D_i|\} \]  \hspace{1cm} (4.5)

In order to develop a method that will help us to find the optimal weights for each case of \( u_i^m \), we assume that the agent selects its weight, \( w_i \), before the opinion sharing round, \( m \), and that it is fixed till the end of the round. If this condition is met, the form of the aggregation function (Equation 3.1) is such that it returns the same result regardless of the ordering of its update sequence (as shown in Figure 4.3, the positions of conflicting updates of the agent’s belief overlap). This implies that the position of the agent’s belief when the strongest support is observed does not depend on the preceding dynamics, and the agent’s weight and prior are the only parameters that regulate this belief position. If the agent can predict the value of the strongest support, \( u_i^m \), that it will observe in the upcoming round, then it needs to consider only 2 weights to form the most accurate opinion given its local view. Specifically, the weight \( w_i^- \) at which the agent’s belief reaches the lower confidence bound \( p_i^k = 1 - \sigma \) to form its opinion \( \alpha_i^k = \text{blue} \) is when the strongest support is observed; or \( w_i^+ \) to reach the upper bound \( p_i^k = \sigma \), to form the opposite opinion. In general, the agent’s prior \( p_i' \) is not equal to 0.5, therefore weights towards different bounds are not equal \( w_i^- \neq w_i^+ \). Since \( u_i^m \in \{1 \ldots |D_i|\} \), we build the corresponding sets of weights:

\[
W_i^- = \{w_i^l^- : l = 1 \ldots |D_i|\} \hspace{1cm} (4.6)
\]

\[
W_i^+ = \{w_i^l^+ : l = 1 \ldots |D_i|\} \hspace{1cm} (4.7)
\]

\[
W_i = W_i^- \cup W_i^+ \hspace{1cm} (4.8)
\]

where \( W_i \) is a set of the candidate weights that the agent needs to consider in order to select the best weight and form the most accurate opinion. Also, this is a complete set of the distinct dynamics of the agent’s opinion formation.

In more detail, we present Algorithm 3 that pre-calculates the candidate weights, \( W_i = \{w_i^l : l = 1 \ldots 2|D_i|\} \), and thus, it heavily reduces the search space from the continuous interval \( w_i \in [0.5, 1] \) to the optimal set \( W_i \).

As we mentioned in the model definition, our approach does not rely on the fact that the model operates with a binary subject of common interest (i.e. \(|B| \neq 2\)). This assumption helps us to simplify the notation. However, the algorithm can be extended for \(|B| > 2\). In particular, in lines 3 and 4 in Algorithm 3 we calculate the candidate weights towards two confidence bounds \( \sigma \) and \( 1 - \sigma \) which represent two conflicting opinions. By changing our notation to express an agent’s beliefs towards a large number of alternatives, we correspondingly increase a number of the confidence bounds. Thus, to adopt our algorithm we need to repeat the same calculations as provided in lines 3 and 4 for these new confidence bounds. Accordingly, we have to update the estimator of the awareness rate defined in Equation 4.10. We should note that by increasing the
Algorithm 3 AAT: Generation of the Candidate Weights

Function CANDIDATEWEIGHTS($p'_i, \sigma, |D_i|$
{Builds a vector of candidate weights}

1: $P(w, u) = \begin{cases} p'_i & \text{if } u = 0 \\ \frac{tP(w, u-1)}{(1-t)(1-P(w, u-1)) + tP(w, u-1)} & \text{otherwise} \end{cases}$
{recursive aggregation function, where $w$ is a weight, $u$ is a number of updates (following Equation 3.2)}
2: $U' := \{1, \ldots, |D_i|\}$ {the number of updates to consider}
3: $W^+_i := \{w^+_i : \text{Solve } (P(w^+_i, u^+_i) = \sigma) \quad \forall u^+_i \in U'\}$
4: $W^-_i := \{w^-_i : \text{Solve } (P(1-w^-_i, u^-_i) = 1-\sigma) \quad \forall u^-_i \in U'\}$
5: $W_i = W^+_i \cup W^-_i$
6: return $W_i$

number of possible states of the subject of common interest, we increase the complexity of the problem and it is therefore likely that AAT will converge more slowly to the critical weights.

However, the reduction of the search space to a smaller set of candidate weights makes the algorithm more efficient. This decreases the number of opinion sharing rounds required for the algorithm to converge to the optimal set of weights. Nevertheless, each candidate still has to be evaluated in order to estimate its awareness rate. The method we offered earlier is computationally expensive since it requires the simulation of each of the candidates individually. In the next section we address this shortcoming.

4.2.2 Heuristic Estimation of the Awareness Rates

In order to select the weight to attribute to the opinions of its neighbours, the agent needs to estimate the awareness rates, $h(w^+_i)$, that would be achieved by using each candidate $w^+_i$. Since the agents’ opinions are highly interdependent, the choice of an individual agent eventually affects the dynamic of the whole system. Therefore, the awareness rate can only be estimated empirically through a number of opinion sharing rounds. However, according to the definition of the awareness rate, $h_i$ (Equation 4.2), the agent can measure it only for the weight, $w_i$, that it currently uses, since there is no direct relation between $h_i$ and $w_i$. By analysing the process of the agents’ belief updating, we propose the following approach to construct an estimator of the awareness rate, $\hat{h}(w'_i)$, for the other candidate weights $w'_i \in W_i \setminus w_i$ based on the observed local dynamics.

Specifically, to estimate the awareness rate the agent needs to decide if its opinion could be formed using a weight, $w'_i$, distinct from the weight it actually uses, $w_i$. We identify two pieces of evidence that indicate that the agent could have formed an opinion:
1. Consider the case that the agent used weight $w_i$ in round $m$ and an opinion was formed ($o_{im}^{} \neq \text{undetermined}$). According to both types of our aggregation functions, (Equations 3.2 and 3.3), all higher weights ($w_l \geq w_i$) would have led to a stronger belief and thus, to opinion formation as well. We formalise this evidence of opinion formation as a boolean function that returns \textbf{True} if the agent would have formed an opinion with a candidate weight, $w_l$, or \textbf{False} otherwise:

$$\text{Ev}_1(w_l, w_i, o_{im}^{}) = (o_{im}^{} \neq \text{undetermined}) \land (w_l \geq w_i) \quad (4.9)$$

2. Otherwise, the opinion should have been formed when the strongest observed support, $u_{im}^{}$, is larger than it is required to cross the nearest confidence bound $\sigma$ or $1-\sigma$ using a candidate weight $w_l$, denoted as $u(w_l, p'_i, \sigma)$. Additionally, we exclude the current weight, $w_i$, which can be more accurately judged by the first piece of evidence. This formulates the second piece of evidence, which is formalised as follows:

$$\text{Ev}_2(w_l, w_i, u_{im}^{}) = (u(w_l, p'_i, \sigma) \leq u_{im}^{}) \land (w_l \neq w_i) \quad (4.10)$$

Combining these two perspectives, we construct an indicator that returns \textbf{True} if the agent might have formed an opinion on the current round, $m$, using weight $w_l$ with the actually used weight $w_i$, or \textbf{False} otherwise:

$$\text{Evs}(w_l, w_i, m) = \text{Ev}_1(w_l, w_i, o_{im}^{}) \lor \text{Ev}_2(w_l, w_i, u_{im}^{}) \quad (4.11)$$

Following the definition of the agents’ awareness rate (Equation 4.2), we formulate the empirical estimator of the awareness rate for each weight out of the candidates $w_l^{} \in W_i$ after the number of opinion sharing rounds $|M|:

$$\hat{h}(w_l) = \left| \left\{ m \in M : \text{Evs}(w_l, w_i, m) = \text{True} \right\} \right| / |M| \quad (4.12)$$

In more detail, Algorithm 4 describes this estimator of the improved version of our algorithm, iAAT, that is executed after each round. In lines 4-10, iAAT updates the estimates of the awareness rate for each of the candidate weights according to the procedure described above. If no opinions were observed ($u_{im}^{} = 0$), the agent cannot form its own opinion with any of the weights and thus this case is limited by the condition on lines 1-3.

Now, following the optimisation problem the agent solves (Equation 4.3), it has to select the weight (line 11) that delivers the awareness rate closest to the target, $h_{trg}$, considering the high interdependence between agents’ choices.

Having defined the AAT algorithm, and its improved version iAAT, we now investigate its properties and choose parameters before evaluating our algorithms and comparing them to the benchmarks.
Algorithm 4 iAAT: Awareness Estimation Rules

Procedure UPDATE(i)
{Revises the current weight after each round}
1: if \( u_i^m = 0 \) then
2: return \{no changes if new opinions did not arrive\}
3: end if
4: for \( l \in \{1, \ldots, 2|D_i|\} \) do
5: if \( \text{EvS}(w_l^i, w_i, m) = \text{True} \) then
6: \( \hat{h}^m(w_l^i) := \frac{m-1}{m} \hat{h}^{m-1}(w_l^i) + \frac{1}{m} \) \{add 1 to the running average\}
7: else
8: \( \hat{h}^m(w_l^i) := \frac{m-1}{m} \hat{h}^{m-1}(w_l^i) \) \{else add 0\}
9: end if
10: end for
11: \( w_i := \text{ChooseWeight}(i) \)

4.3 Analysis of AAT Parameters

Our algorithms have a single parameter, the target awareness rate \( h_{\text{trg}} \), that each agent aims to achieve. Since iAAT is an extension of AAT designed to reduce computational cost, in this section we focus on analysis of the more generic version of our algorithm, AAT.

We have already identified in Section 4.1.1, that accuracy \( R \) is maximised when agents use the minimal weight that still results in a high awareness rate. Specifically, when the awareness rates are slightly lower than the maximum, this indicates the turning point in the dynamics of the sharing processes when accuracy is improved. However, in this area of the optimised parameter settings, the system does not always disseminate information on a large scale and as a result the awareness rates of the agents may suffer even further. Given this, in Section 4.3.1 we validate the intuition on which AAT is built, by evaluating the influence of the target awareness rate on the accuracy of consensus.

After selecting the target awareness rate to use in our experimental setup, we study in Section 4.3.2 the performance of different AAT strategies that select a weight out of the candidates. We aim to examine our hypothesis that the strategies which introduce less dramatic changes in system dynamics enable better estimation of the awareness rates of the candidate weights and thus, achieve higher performance.
Chapter 4 Accurate Consensus with Anonymous Peers

Figure 4.4: The accuracy of the system of $N = 1000$ agents depending on the selection of the target awareness rate $h_{trg}$ (left column) and the average weight achieved by AAT with the corresponding given $h_{trg}$ (right column). Each data point represents a result averaged over 5 experiments. Results for $\langle d \rangle = 6, 10$ are skipped for brevity, but included in the average (f).
4.3.1 Selection of the Target Awareness Rate

We analyse the performance of our algorithms with regards to its single parameter, the target awareness rate, $h_{trg}$. Among different AAT strategies for selecting the weight, for this experiment we chose the hill-climbing strategy and will look at the performance of other strategies later. Note that all our strategies use the same approach (Algorithm 4) to estimate the awareness rates and thus, the choice of a specific target awareness rate will have the same qualitative effect on all strategies. Therefore, the results of the following experiment might be applied to other strategies, such as $\epsilon$-greedy, $\epsilon$-N-greedy, greedy and soft-max strategies, which has been confirmed by our additional studies.

The value of the target awareness rate $h_{trg}$, when accuracy $R$ is maximised, depends on a number of the model parameters that influence the dynamics of the opinion sharing process. We discussed them in choosing the experimental setup in Section 3.2.2. In particular, we identified that the properties of the communication network are the most influential on the opinion sharing processes and thus, are the most relevant to examine compliance with the research requirements. Therefore, in this experiment we evaluate systems of $N = 1000$ agents simulated on networks produced by three different topology generators selected for our experiments. In this setup, we investigate the influence of the $h_{trg}$ on the algorithm's performance.

Our empirical study shows that the target awareness rate that delivers the highest accuracy is maximised around a single value over a number of different network parameters. Specifically, Figure 4.4 shows that the system exhibits the highest accuracy when $h_{trg}$ is close to 0.9 for different topologies and expected degrees. We present results, along with the average weight (defined as $\langle w_i \rangle = \frac{1}{N} \sum_{i \in A} w_i$) that AAT selects with a given $h_{trg}$, that help to explain the significant drop in accuracy for the higher values of the target awareness rate. As can be seen, when $h_{trg} > 0.9$ the agents select significantly larger weights to form opinions out of a smaller number of observations. Thus, the agents become overconfident and the whole system converges to consensus quicker, with the distributed aggregation process becoming less distinct on a system scale.

Based on the results of this experiment, in our further evaluation we use AAT with $h_{trg} = 0.9$. The selection strategy of AAT is responsible for the agents reaching this target awareness rate, but as we have argued above, it does not influence the value of $h_{trg}$ when the accuracy is optimised, since its value depends on the model parameters. However, each selection strategy has its own effects on the dynamics of weight tuning and, as a result, on the achieved accuracy. In the next section we analyse their differences.

4.3.2 Comparison of AAT Weight Selection Strategies

In this section we test our hypothesis that AAT based on the hill-climbing strategy converges to better parameters that result in higher accuracy. Specifically, we assumed
that a strategy that introduces less sudden change to the opinion sharing process will estimate the awareness rates more accurately. In order to test this hypothesis, we evaluate the accuracy reached by different strategies and analyse their weight tuning dynamics.

![Accuracy of consensus, R%](image)

**Figure 4.5:** The accuracy of the system of \( N = 1000 \) agents depending on the selection of strategy for AAT compared with the existing solution DACOR. Each data point represents an averaged result over 5 experiments with variable expected degree \( \langle d \rangle = \{4, 6, 8, 10, 12\} \).

To this end, Figure 4.5 presents the accuracy reached by each strategy in this experiment with an additional comparison against the state-of-the-art algorithm, DACOR. We can see that the \( \epsilon \)-greedy strategy exhibits the worst performance since it introduces a large number of sudden changes to the opinion sharing process. This hypothesis is confirmed by studying the corresponding dynamics of weight tuning over a number of opinion sharing rounds presented in Figure 4.6. Thus, due to the high interdependence in the system, this strategy is not able to estimate the awareness rates of the candidate weights and converge to the solution. Similar results are shown by the \( \epsilon \)-N-greedy strategy with a slight improvement since its randomness decays with time. Despite its simplicity, the greedy strategy forces agents to keep a previously selected weight for a longer period and with more stable dynamics the system converges to a much better solution. However, a large number of agents may change their weights simultaneously and thus, AAT may fluctuate around the optimised weights. The soft-max strategy provides better results by selecting the weight with the awareness rate closest to the target with a higher probability. Finally, the hill-climbing strategy introduces the least changes, estimates the awareness rate of the candidate weights with the most accuracy and, as a result, exhibits the highest accuracy overall. This experiment shows that the latter three strategies all outperform the results achieved by the existing state-of-the-art algorithm, DACOR, and in the following experiment we analyse this in much wider settings. Additionally, it confirms our earlier hypothesis that the algorithms are sensitive to topological properties. Specifically, the small-world topology leads to a significantly
different performance. Therefore, in the following experiments we study performance of the algorithms on each type of network topology.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.6.png}
\caption{The sample dynamics of the average weight for different AAT strategies. Results for a same system of $N = 1000$ agents with a random communication network with $\langle d \rangle = 8$.}
\end{figure}

In summary, the empirical study of AAT confirmed our hypothesis that the hill-climbing strategy delivers the highest accuracy. Also, we analysed the influence of a value of the target awareness rate on accuracy and we found that, for the experimental setup, the optimal target awareness rate is close to the 0.9. Considering these results, in the following experiments we use AAT based on the hill-climbing strategy with $h_{\text{trg}} = 0.9$.

\section{4.4 Empirical Evaluation}

In this section we empirically evaluate our algorithm AAT and its improved version iAAT. In so doing, we investigate the properties of AAT and examine its compliance with the research requirements introduced in Chapter 1. Despite AAT being the first algorithm that improves the accuracy of consensus based only on agents’ local views, we benchmark it against the state-of-the-art algorithm, DACOR, which improves the accuracy of consensus by exchanging service messages to find the optimised parameters. Both algorithms pursue the same goal of self-organising a system into a parameter setting in which the accuracy of consensus is significantly improved. This gives us the ability to perform a direct comparison of their relative performance. Additionally, we compare the performance of the algorithms with the benchmarks based on the centralised pre-tuning of a system which was introduced in Section 3.5.
Table 4.1: Experimental setups for the algorithms evaluation

<table>
<thead>
<tr>
<th>Model parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agents’ aggregation function</td>
<td>$f(\cdots)$</td>
<td>{Bayesian, Weighted sum}</td>
</tr>
<tr>
<td>Number of agents</td>
<td>$N$</td>
<td>{100 . . . 10000}</td>
</tr>
<tr>
<td>Network topology</td>
<td>-</td>
<td>{Random, Scale-free, Small-world}</td>
</tr>
<tr>
<td>Expected degree</td>
<td>$\langle d \rangle$</td>
<td>{8, 100}</td>
</tr>
<tr>
<td>Fixed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agents’ priors</td>
<td>$p'_i$</td>
<td>drawn from $\mathcal{N}(\mu = 0.5, s = 0.09)$</td>
</tr>
<tr>
<td>Agents’ confidence bounds</td>
<td>$(1 - \sigma, \sigma)$</td>
<td>(0.2, 0.8)</td>
</tr>
<tr>
<td>Number of sensing agents</td>
<td>$N_s$</td>
<td>0.05 $\cdot$ $N$</td>
</tr>
<tr>
<td>Accuracy of introduced opinions</td>
<td>$r$</td>
<td>65%</td>
</tr>
<tr>
<td>Rate of opinion introduction</td>
<td>$\lambda$</td>
<td>every 10 steps</td>
</tr>
<tr>
<td>Number of introduced opinions</td>
<td>$\Lambda$</td>
<td>$3 \cdot N_s$</td>
</tr>
<tr>
<td>Number of opinion sharing rounds</td>
<td>$</td>
<td>M</td>
</tr>
</tbody>
</table>

Since one of our motivations for developing an algorithm for accuracy improvement results from the difficulties of theoretical analysis, we conduct an empirical study. Specifically in Section 3.4, we showed that the critical mode is very sensitive to initial parameters. Additionally, we discussed that the model cannot be simplified in order to enable its analytical analysis without losing its properties. Therefore, in order to investigate the applicability of our solution in a variety of realistic settings that cannot be analytically analysed, we evaluate AAT empirically on the experimental setup offered earlier in Section 3.2.2. The summarised version for this set of experiments is presented in Table 4.1.

In the following subsections we analyse the metrics offered in the model definition. Specifically, we study the accuracy of consensus achieved by the algorithms in Section 4.4.1, their communication expenses in Section 4.4.2 and their computational expenses in Section 4.4.3. Next, we examine the robustness of the algorithms in Section 4.4.4. In particular, we investigate the accuracy of a system with a number of indifferent agents that do not participate in the optimisation process.

### 4.4.1 Accuracy of Consensus

In this section our aim is to examine AAT and iAAT in terms of our main research objective. Specifically, we measure the accuracy of consensus, $R$, defined in Equation 3.6, in variable settings of our experimental setup in order to examine the compliance of AAT with the research requirements. We compare our algorithms with the benchmarks,
DACOR\(^1\) and the three cases of the centrally pre-tuned system, introduced Section in 3.5. Specifically, the later benchmarks indicate the level of accuracy that can be achieved if it would be possible to: (i) find the critical weight for each instance of the system individually by computationally intensive empirical exploration of possible weights; (ii) predict the critical weight by using the average value identified for systems with the same topological properties (size, degree and network topology); and (iii) select weights such that the system operates in the unstable mode. None of them are likely in practice, however, together with the theoretical optimum, they constitute bounds on possible performance.

In more detail, the results of the accuracy benchmark are shown in Figure 4.7 and the rest of the outstanding experiments can be found in Appendix B (for networks with the average degree \(\langle d \rangle = 100\) and agents based on the weighted sum aggregation function). As can be seen, AAT and iAAT exhibit similar performance and significantly outperform DACOR. Despite tuning for our model, DACOR cannot improve the accuracy of consensus. It self-organises the system into the unstable mode and delivers a level of accuracy similar to the system when pre-tuned into the unstable mode and the analytically-predicted accuracy of this mode \(R_{\text{min}2}\) (see Equation 3.11). This result is another confirmation that the branching factor, which DACOR computes, is not a reliable indicator of the critical mode in our model.

In contrast, the accuracy of consensus achieved by AAT and iAAT is close to the results of the individually pre-tuned system despite being lower than the theoretical maximum \(R_{\text{max}}\) for a centralised system of the same size. Detailed analysis of the results shows that the performance of DACOR is highly dependent on its parameters, which have to be individually tuned for a specific domain and thus, on average DACOR delivers low accuracy. In contrast, AAT and iAAT exhibit equally high adaptivity to the variety of the settings we considered in this experiment. Crucially, improvements we introduced into iAAT design do not harm its performance.

Despite the fact that the individually pre-tuned systems exhibit higher accuracy than AAT, we noted in the definition of these benchmarks that they are computationally extremely expensive and cannot be applied in realistic settings. This also limits the maximum size of a system we can evaluate in this experiment. However, this benchmark provides us with an insight of the upper bound of the accuracy that can be reached by tuning the opinion sharing processes in a system. Also, the accuracy of the systems evaluated with the average critical weight, \(\langle w_c \rangle\), confirms that the value of the critical weight is sensitive to system parameters, such as specific shape of its network topology. Ultimately, this result indicates that the average critical weight is not a reliable approach
for reaching the critical mode even when all system parameters are known. This confirms the need for an adaptive algorithm that can reliably improve the accuracy of any system.

Lastly, we can conclude from this experiment that AAT scales well, since it reaches the stable level of accuracy around 80-88% for systems larger than 1000 agents on all tested topologies. However, accuracy declines as the system size becomes lower than 1000 agents, since all approaches rely on the properties of collective behaviour in the model. Collective behaviour is less distinct in smaller systems and, therefore, AAT and the other benchmarks deliver lower accuracy. This was expected following our analysis of the model performance in the previous chapter.

4.4.2 Communication Expense

AAT is designed to improve the accuracy of a system without introducing additional communication above opinion sharing as described by the model. However, the agents

Figure 4.7: Accuracy of consensus achieved by the algorithms and the benchmarks depending on the system size and topology ($\langle d \rangle = 8$, Bayesian aggregation function)
still have to communicate in order to share their opinions and to filter the inaccurate ones in a distributed manner. In this section we compare the number of messages that agents exchange in order to find their opinions (i) while the system is tuned by AAT and iAAT, (ii) with the total number of messages including service messages required to operate for DACOR and (iii) with the minimal communication, $U_{\text{min}}$ (Equation 3.12), defined as the number of messages required to share an opinion on a system scale in a single opinion cascade. The comparison against the latter benchmark shows how much communication is introduced above that of the bare minimum required for agents to form their opinions.

In more detail, Figure 4.8 presents the average number of messages exchanged in a system per opinion sharing round against the system size, where results are averaged across all the system instances we evaluated in the previous experiment. As we discussed earlier, DACOR requires a significant communication overhead in order to optimise the opinion sharing process. In contrast, AAT does not introduce additional communication and, even for systems with a large number of agents, the communication overhead required to improve the accuracy is not notable. The average number of messages for a system with AAT is the same as the minimal communication, because during some rounds a system with AAT does not disseminate opinions on a large scale (as a result of the fact that the target awareness rate in AAT is lower than the maximum, $h_{\text{trg}} < 1$). In this metric iAAT again exhibits very similar behaviour.

The results indicate that AAT operates in the area in which communication is close to the minimal, whilst it also significantly improves the accuracy of consensus. Thus, our algorithms meet the research requirement of communication efficiency. Moreover, the results confirm the scalability of our algorithms since the number of messages exchanged in a system coincides with minimal communication. Considering the definition of minimal communication, any further reduction of communication is impossible without harming the accuracy of the system, since some of the agents would not be able to form opinions.

### 4.4.3 Computational Expense

Now, to investigate the efficiency of our solution we evaluate the computational costs it introduces in order to improve accuracy. The complexity of a single run of our algorithm, AAT, and the benchmark DACOR, is insignificant compared to the total number of runs required in the process of weight tuning. Therefore, we measure the computational cost as: i) the number of times the algorithms are changing the agents’ weights during an opinion sharing round; ii) the time required to simulate the system.

To this end, our results presented in Figure 4.9 show that AAT introduces radically fewer changes of the agents’ weights ($4 \cdot 10^4$ times less changes per agent in a system of 1000 agents with the expected degree $\langle d \rangle = 8$) in the process of finding the optimised
parameters, than DACOR. More specifically, AAT and iAAT update the weight of an agent only once at the end of each round, while DACOR updates an agent’s weight if any of its neighbours has observed a new opinion.

Considering computational cost as the time required to simulate the system, we compare such expenses of the algorithms in Figure 4.10 with the simulation of the systems with fixed weights (the “Individually tuned $w_c$” benchmark). This comparison provides us with a base line and shows that the computational cost of running DACOR and iAAT is close to the cost of simulating the system without behavioural algorithms. However, the computation cost of AAT is significantly higher and thus, it might be too expensive to deploy it in agents with limited resources. Such performance confirms our earlier hypothesis that some stages of AAT impose high computational cost and justifies the development of the iAAT extensions in Section 4.2. Considering the previous metrics, it is notable that the computational cost is the only significant difference between AAT and iAAT versions of our behavioural algorithm.

Additionally to the computational cost, we should consider the memory requirements of each algorithm. DACOR (Algorithm 1) requires for each agent to store only $\Delta \alpha'_i$, 

![Figure 4.8: Communication expenses for AAT, iAAT and DACOR depending on the size of the system. Error bars are not noticeable on the scale of the plots. Communication expenses for AAT and iAAT overlap together with the minimal communication, which is required to share a single opinion between all agents.](image)
which is the previous value of the local branching factor. AAT algorithm requires to initialise its search space with the set of the candidate weights, $W_i$, each assigned its own awareness rate ($\hat{h}(w^l) \forall w^l \in W_i$). Apart of this, during each opinion sharing round AAT records a history of received opinions, and uses it to update the awareness rates of all candidate weights. In contrast, iAAT does not record the history of received opinions. Instead, iAAT stores only the strongest observed support, $u^m_i$. Moreover, the number of candidate weights required by iAAT is limited to the optimal set, which in our simulation is on average a magnitude smaller than sets used by AAT. Thus, iAAT requires to store in an agent’s memory a table to candidate weights and their awareness rates (10...100 values in total), and a single dynamic value, $u^m_i$.

So far we have demonstrated that AAT meets the research requirements of delivering high accuracy, adaptivity, scalability, and communicational and computational efficiency with the iAAT extensions. In the next section we examine the last requirement of solution robustness.

4.4.4 Heterogeneous Agent Population

In this section we consider heterogeneous systems in which some of the agents do not participate in the optimisation process. By doing so, we examine the robustness of the AAT algorithm in settings where not all agents can be curated by the same behavioural algorithm.

More specifically, in large systems it might be infeasible to deploy a tuning algorithm simultaneously on all agents. A number of agents might have very limited resources to extend their functionality or some agents might be faulty. In order to investigate
performance in such settings, we evaluate the accuracy that the algorithms can reach in systems where a number of agents are indifferent and their weights are not dynamically determined by the behavioural algorithms. We simulate these settings by introducing a number of indifferent agents that are randomly distributed across the system. The weight used by each indifferent agent is fixed and uniformly drawn from a range of [0.55, 0.75], which is widely distributed around the critical weight (see Figure 3.10B, as an example of a system evaluated in our experimental setup).

The results of this experiment with a system of 1000 agents, the expected degree $\langle d \rangle = 8$ and a variable share of indifferent agents are presented in Figure 4.11. As can be seen, with up to 80-90% of indifferent agents, with fixed weights, AAT delivers a higher accuracy than the accuracy of introduced opinions $R_{\text{min}2}$. This shows the direct benefit from deploying AAT even on a fraction of the agents in the system.

AAT exhibits high performance in such settings because the actions of the agents depend only on locally available information, and thus, independent from actions chosen by other agents (or lack of actions in the case with indifferent agents). Crucially, that as the number of indifferent agents increases, the agents running AAT are still able to
Chapter 4 Accurate Consensus with Anonymous Peers

Figure 4.11: The accuracy of a system with AAT in a heterogeneous agent population of a scale-free network (results for other topologies are similar)

efficiently affect the system dynamics. At the same time, the performance of DACOR confirms our earlier claims that it is not optimised for our settings.

In this experiment we have showed that AAT improves the accuracy of a system even when only installed on a fraction of the agents in a system. Specifically, our behavioural algorithm is the first solution that significantly improves accuracy when up to 80-90% of the agents do not participate in the optimisation process. This indicates high robustness and thus, AAT meets the corresponding research requirement.

4.5 Summary

In this chapter, we presented the Adaptive Autonomous Tuning algorithm which is the first solution to improve the accuracy of consensus in large decentralised systems without introducing any communication beyond basic opinion sharing. A system tuned with the agents’ behaviour curated by our AAT algorithm operates in the critical mode of the opinion sharing process, which implements a decentralised opinion aggregation. We developed AAT relying on the insight that the critical mode is induced when the weights of the agents are minimally sufficient to share their opinions. This creates conditions in which early and possibly inaccurate opinions are shared in smaller groups to prevent overreaction. Only when groups with the same opinion merge is this locally-supported opinion disseminated on a larger scale. To find these conditions, AAT helps each agent to individually select a minimal weight that still leads to its opinion formation. In particular, we described the three main stages of this process: (i) to form a set of the candidate weights; (ii) to estimate if each of the candidate weights would have led to the
agent forming an opinion in the current round; and (iii) to select the minimal weight out of the candidates that form the agents’ opinion with the target awareness rate. The latter stage implies that the agents have to compromise their awareness rates in order to achieve the area of optimised parameters. Finally, we showed that some of the stages can be improved upon by relying on additional information and proposed the iAAT algorithm, which is computationally more efficient.

We empirically evaluated our algorithm in order to investigate its properties. As a result, we showed that AAT meets the research requirements identified in Section 1.4 by delivering high levels of:

1. **Accuracy** of the agents’ opinions and thus, high accuracy of consensus. Specifically, we benchmarked AAT against the current state-of-the-art algorithm, DACOR, and against a pre-tuned system for the highest accuracy. We showed that AAT significantly outperforms DACOR and delivers accuracy close to that of a system pre-tuned by empirical exploration. However, the pre-tuning of a system requires centralised coordination and large amounts of computational resources to find the optimised parameters. Thus, AAT is currently the best solution to improve the accuracy of consensus in large decentralised systems.

2. **Adaptivity**: We showed the high adaptivity of AAT by evaluating it on a number of different network topologies with variable densities. Additionally, we demonstrated that it is impossible to predict the critical weight which improves accuracy by analysing the best parameters of pre-tuned systems. Thus, we confirmed the need to develop an adaptive approach that improves the accuracy of each system individually. At the same time, we identified low adaptability as a significant weakness of DACOR, which requires tuning of its parameters in order to achieve the highest performance in specific settings.

3. **Scalability**: We evaluated the algorithm on systems with up to 10000 agents and showed that the solution scales well. In particular, our approach exploits collective behaviour, which explains why AAT delivers higher accuracy in larger systems. At the same time, the computational cost for each individual agent remains constant. As a result, we concluded that it can be used in much larger systems that were not simulated due to the high computational expenses of the pre-tuned benchmarks.

4. **Communication efficiency**: This requirement was met in the design of the algorithm, which does not introduce additional communication over that already present opinion sharing without any supporting information. We demonstrated that the communication exchange in the system curated by AAT is within the range of error of the minimally required communication to share the opinions. Conversely, DACOR introduces additional service messages and communication expenses are therefore significantly higher.
5. **Computational efficiency**: We showed that, unlike DACOR, AAT requires significantly fewer weight changes to reach the critical mode. However, at the same time AAT is computationally expensive. Its improved version, iAAT, has similar cost to that of DACOR, and importantly, iAAT is close to the cost of running a system without any behavioural algorithm. Finally, in order to operate, AAT requires to store hundreds of variables in memory, while iAAT several tens.

6. **Robustness**: We demonstrated that AAT can be deployed in heterogeneous systems which include agents that do not participate in the optimisation process. The improvement of the accuracy of consensus drops linearly with the number of such agents. Notably, that even with 10-20% of agents running AAT, the accuracy of consensus significantly improves in comparison to the accuracy of introduced opinions. Thus, AAT is highly robust and it can be deployed in highly heterogeneous systems.

By meeting the listed requirements above, our algorithm is the first solution that improves the accuracy of a system with minimal communication requirements. It also outperforms the existing solution and meets our research requirements in the more difficult case in which the peers are anonymous.

More specifically, the AAT algorithm attributes a common weight to all the network neighbours of an agent, assuming that it cannot differentiate its peers. As we identified in our research requirements, such a limitation is essential to apply our solution to anonymous networks. However, we should also consider the case in which an agent can identify its peers and benefit from this knowledge. Therefore, in the following chapter we approach this outstanding research problem.
Chapter 5

Accurate Consensus with Identified Peers

In this chapter we tackle our last research aim. In the previous chapter we considered the case where the communication network is anonymous and agents cannot differentiate between their network neighbours. However, as we identified in our motivating scenario, in many cases agents can observe their neighbourhood, and thus could potentially benefit from identifying the sources of received opinions. For example, the underlying communication protocol may require senders to identify themselves or agents may operate in a wired network where each communication channel is dedicated to a specific pair of peers.

In the scenario with identified peers, which we introduced in Chapter 1 as Requirement 2b and consider in this chapter, every received opinion is annotated with its sender. This generates more information for an agent’s behavioural algorithm, and opens a challenge as to how this annotation should be used in order to improve the accuracy of consensus.

In terms of our model, the new behavioural algorithm is able to differentiate between the network neighbours, namely peers, and can therefore attribute different weights to their opinions. This is a crucial difference from the AAT algorithm designed in the previous chapter. Apart from the problem of assigning weights individually to each peer, the significantly larger number of variable weights also poses a new challenge of timely convergence to a state of the system when the accuracy of consensus is improved. The latter challenge is introduced by the significantly larger search space the new algorithm has to analyse and the more challenging experimental setup we intend to focus on. Specifically, identification of peers is more important in dense communication networks, and in this chapter we focus on systems with high expected degree of their communication networks.
In the following section we analyse in greater detail the problem of determining preferences among identified peers, and look into the performance of the model and the AAT algorithm in dense networks. This section outlines how AAT can be improved by identifying important peers. In order to solve this problem, we need to analyse how agents should: (i) determine their preferences among the peers, and then (ii) attribute the individual weights to them in order to induce the desired mode of collective behaviour. Following this, in Section 5.2 we identify and analyse indicators of a peer’s relative importance. Then we design extensions of the AAT algorithm which assign a common weight to the most important peers selected according to each indicator. Evaluation of these extensions provides us the most appropriate indicator we will use in the second step. Specifically, in Section 5.3 we present a new behavioural algorithm, the Individual Weight Tuning (IWT) algorithm, which identifies weights for each peer individually. We evaluate and analyse its performance in Section 5.4, and finally, conclude the chapter in Section 5.5.

5.1 The Problem of Determining Preferences among Peers

The problem we set out to solve in this chapter is to improve the accuracy of consensus even further than we achieved in Chapter 4. In comparison to the previous scenario, the only additional feature of an agent is its ability to identify the sender of an opinion. Clearly, this may assist in making a more informed decision and, ultimately, in forming a more accurate opinion. In this section we analyse aspects of our new problem and, specifically, how preferences among identified peers can be determined.

In the following subsection we analyse the increase in the complexity of our problem. Then, in Subsection 5.1.2, we discuss in detail why the identification of peers is expected to bring benefits only in dense networks. Following this system overview, we analyse a single agent and its local view in Section 5.1.3. This analysis provides us with a number of possible inputs for behavioural algorithms. Finally, in Section 5.1.4 we outline an approach to developing the required behavioural algorithm.

5.1.1 The Number of Variable Weights in the System

To solve our research problem in the challenging circumstances in which communication is strictly limited to opinion sharing, we rely on the properties of collective behaviour in the opinion sharing process. More specifically, we looked for the parameters that induce the critical mode of collective behaviour in which a system self-organises into a distributed filter thereby significantly increasing the accuracy of consensus. To achieve this, we gradually increased the complexity of our solution. Firstly, in Chapter 3, we evaluated the model with a fixed weight common to all agents. This simplification based
on centralised weight tuning, enabled us to analyse the model’s behaviour and design the pre-tuned benchmarks. Then, in Chapter 4 we designed the first behavioural algorithm, AAT, in which each agent is independently able to choose the weights it attributes to other agents. Now, we approach the problem of designing a new behavioural algorithm, the Individual Weights Tuning (IWT) algorithm, which differentiates between agents’ neighbours and attributes individual weights to their opinions.

The increasing complexity of our solutions, as the number of weights required to determine, is summarised in Table 5.1. In contrast to \( N \) weights that agents running AAT tune in the system, the new solution, IWT, has to tune \( N \cdot \langle d \rangle \) weights, where \( \langle d \rangle \) is the expected degree of the communication network of \( N \) agents. Due to the significantly larger number of variable weights in the system, the new algorithm faces a risk of slow convergence to the global solution which induces the critical mode of behaviour.

This risk is particularly apparent if preferences over peers are made in dense networks with \( \langle d \rangle \geq 50 \). In this case, the control parameter for the complexity of the IWT algorithm, \( \langle d \rangle \), becomes a significant factor. In the following subsection we discuss why dense networks are particularly promising for applying the IWT algorithm.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Number of Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individually pre-tuned system (benchmark)</td>
<td>( {w_c} = 1 ) fixed weight for all agents</td>
</tr>
<tr>
<td>Autonomous Adaptive Tuning (AAT) algorithm</td>
<td>( {w_i : i \in A} = N ) each agent chooses a weight which it attributes to all its peers</td>
</tr>
<tr>
<td>Individual Weights Tuning (IWT) algorithm</td>
<td>( {w_{ij} : j \in D_i, \ i \in A} = N \cdot \langle d \rangle ) each agent chooses weights individually for each of its peers</td>
</tr>
</tbody>
</table>

5.1.2 Opinion Sharing in Dense Communication Networks

In the previous chapter we showed that AAT is especially efficient for accuracy improvement in systems with sparse communication networks, \( \langle d \rangle = 8 \), where it performs very close to our pre-tuned benchmarks (Section 4.4.1). At the same time, in systems with dense communication networks, \( \langle d \rangle = 100 \), AAT does not follow our benchmarks that closely (see Figure B.1 in Appendix B), which potentially leaves 15% available for accuracy improvement.

This weakness of the algorithm comes from its design, which we discussed above. Specifically, the agents controlled by AAT attribute the same weight to all their neighbours. In the critical mode of a system with a sparse network, agents receive on average less
than two opinions from their peers before forming their own opinion. Since this number is relatively low, the agents do not have a choice over their opinion sources. Thus determining individual preferences among peers is unlikely to bring noticeable benefits in improving the accuracy of consensus.

In contrast, in the critical mode of a system with a dense communication network, agents receive a significantly larger number of opinions from their peers before forming their own opinion. Due to the high connectivity within dense networks, the same opinion is likely to arrive to an agent via multiple opinion sharing paths. If the agent applies the same weight to all its peers, it may suffer from the double counting problem by aggregating an opinion originating from the same sensing agent several times, thereby forming an overconfident belief. Agents that run AAT suffer this negative effect, which explains lower level of the accuracy improvement in dense networks. In order to mitigate this, the IWT algorithm should attribute higher weights to peers that deliver new opinions, and lower weights to peers that communicate the already aggregated opinion and do not contribute to forming a more accurate opinion. However, since the opinions shared are missing any annotation, the agents cannot directly decide if they have aggregated any particular opinion or not. Therefore, we need to design indicators which enable them to deduce such information from local opinion sharing dynamics.

Considering this discussion, we expect IWT to perform better in dense networks, where agents running AAT would suffer from the double counting. In order to make preferences between peers in IWT, we must first identify which information is locally available to an agent for such decision making.

5.1.3 Agents’ Local Views

Before we design new behavioural algorithms, we need to analyse what is available to an agent in its local view considering the updated requirements. Such an analysis provides us with a number of inputs for the behavioural algorithms we design in the following sections.

Following the definition of our opinion sharing model we presented in Chapter 3, each agent has its own private belief, its opinion formed from it and a number of weights it attributes to the opinions of its peers. Additionally to this, an agent is able to observe the dynamics of these variables through time. Considering that weights are the only parameter an agent controls with a behavioural algorithm such as AAT, we focus on others which may bring any additional information.

In more detail, Table 5.2 summarises the list of variables observable by an agent. In our design of the AAT algorithm we have already identified that the awareness metric, which is the probability of an agent forming its opinion, is an indicator of the critical
Chapter 5 Accurate Consensus with Identified Peers

Table 5.2: Local view of agent $i$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>$k \in [1\ldots\infty)$</td>
<td>simulation timestep</td>
</tr>
<tr>
<td>Private belief</td>
<td>$p^k_i \in [0\ldots1]$</td>
<td>the private belief on timestep $k$</td>
</tr>
<tr>
<td>Own opinion</td>
<td>$o_j^k \in {\text{orange, blue, undet.}}$</td>
<td>own opinion on timestep $k$ formed out of its own private belief $p^k_i$</td>
</tr>
<tr>
<td>Received opinion</td>
<td>$o_j^k \in {\text{orange, blue}}$</td>
<td>opinion received from peer $j$ on timestep $k$</td>
</tr>
<tr>
<td>Opinion source</td>
<td>$j \in D_i$</td>
<td>peer $j$ that communicated opinion $o_j^k$ and weight $w_{ij}$ assigned to it</td>
</tr>
</tbody>
</table>

mode of the system (Section 4.1.1). By so doing, we made a link between the local view of a single agent and the state of collective behaviour.

The next step is to design new indicators which, relying on the described local view of an agent, enable it to make preferences between its peers and contribute to the accuracy of consensus. Before discussing such indicators, we must first outline how behavioural algorithms can actually determine preferences among peers.

5.1.4 Behavioural Algorithms for Determining Preferences among Peers

In order to attribute individual weights to its peers, an agent has to solve two problems. First, it has to identify which of its peers’ opinions should be the most influential in the process of its own opinion formation. Second, the agent has to convert these preferences into weights which will induce the critical mode of collective behaviour in its neighbourhood. These two stages are highly interdependent. If an agent modifies its preferences for its peers, this influences local opinion dynamics. In order to return the local dynamics into the critical mode, the agent must then update all the weights it attributes to its peers.

In order to decompose these problems, we split our algorithm design into two stages:

1. The behavioural algorithm which limits connectivity of an agent. It does so by ignoring opinions from less preferable peers. This algorithm requires an additional parameter, the connectivity threshold, which defines when a peer should be ignored. For the rest of its peers, which are selected as the preferable ones, an agent can apply the AAT algorithm without further modifications. This simplified design enables us to analyse different indicators of peer importance with regards to their influence on the accuracy of consensus. At the same time, we avoid solving the problem of converting these preferences into the agents’ weights.

2. The behavioural algorithm which attributes individual weights to agents’ peers is the IWT algorithm discussed above. In particular, once we identify a reliable
indicator of peer importance, we have to solve only the second problem in the algorithm design. We do this by defining the individual weights for peers as a combination of an agent’s preferences and the indicator on the critical mode we developed earlier along with AAT.

In the following two sections we design such algorithms. Specifically, Section 5.2 discusses the first stage in which we design and evaluate four indicators of peer importance. Then, using the chosen indicator, we analyse the second stage in Section 5.3, which presents the IWT algorithm.

## 5.2 Limiting Connectivity According to the Preferences

In this section we design behavioural algorithms which limit the connectivity of an agent. In this analysis our main goal is to develop and evaluate an indicator of the importance of each peer. We offer two indicators based on the local view of an agent and two benchmark indicators. The latter are the random importance of a peer and the importance of the shortest path length to the sensing agents, which requires external knowledge of the network topology.

We evaluate each indicator using an extended version of the AAT algorithm, which connects only to a limited number of peers selected by the corresponding indicator. In these experiments agents apply the same weight to the set of selected peers, ignoring the opinions of the others. The exact number of peers to connect, which is the connectivity threshold \( x \), is the only common parameter for the indicators we evaluate. It becomes an additional variable in the experimental setup, which we used earlier for the model evaluation in Section 3.3.1 and analysis of AAT parameters in Section 4.3. Following our discussion of the influence of network density on the performance of the algorithms in Section 5.1.2, we choose a high expected degree, \( \langle d \rangle = 100 \), which defines the number of peers in the neighbourhood of an agent. In these circumstances we analyse the accuracy of consensus achieved by this extended version of AAT, depending on the connectivity threshold. The results enable us to analyse the performance achieved using each indicator and to conclude which one to use in the IWT algorithm we design later.

Now, we introduce the indicators and evaluate them. Subsection 5.2.1 discusses the baseline benchmark, which randomly assigns the preferences. Another benchmark, based on the intuition that topological properties can be indicative of peer importance, is presented in Subsection 5.2.2. Section 5.2.3 explores how the time of opinion arrival can indicate peer importance. Finally, Section 5.2.4 presents an indicator which measures how surprising are the opinions communicated by each peer.
5.2.1 Randomised Preferences as a Benchmark

In order to set up a benchmark for the following indicators of peer importance, first we investigate the *random strategy* for determining preferences among peers. Using this strategy, an agent ranks peers in its neighbourhood randomly without actually analysing their behaviour. Alongside this, we introduce an extended version of the AAT algorithm and its evaluation approach.

**Algorithm 5** AAT: Limiting the agent’s connectivity following its preferences

```
Procedure AATLimitConnectivity(i, l) {After each round revises the weights w_ij, which agent i attributes to its neighbours j ∈ D_i. }

1: Attribute preferences to each peer, φ_ij, in this case randomly:
   φ_{ij}^{rand} = RANDOM() ∀ j ∈ D_i
2: Form subset E_i of x selected peers with the highest preferences:
   E_i = GetMax(x elements from D_i by φ_{ij}^{rand}); E_i ⊆ D_i, |E_i| = x
3: Ignore opinions from the other peers:
   w_{ij} = 0 : ∀ j /∈ E_i ;
4: Define the common weight to the chose peers using the AAT algorithm:
   w_{ij} = AAT-Update(i) : ∀ j ∈ E_i
```

In more detail, Algorithm 5 describes the necessary extensions of the AAT algorithm. Specifically, we limit the connectivity of an agent from the initial set of \( D_i \) neighbours as defined by network topology, to the set of selected peers, \( E_i \). For this benchmark, we populate \( E_i \) with \( x \) randomly selected peers from \( D_i \) accounting to the indicator of peer importance, \( φ_{ij}^{rand} \), where \( x \) is the connectivity threshold limiting the maximum number of neighbours \( (x ≤ |D_i|) \).

When the connectivity threshold, \( x \), is small, the communication network becomes sparse and agents receive fewer opinions from their peers. Therefore, the better the indicator of peer importance, the smaller the drop in the achieved accuracy of consensus to be expected. To evaluate this, we vary \( x \) in the range from 1 to the highest degree in the network when \( E_i = D_i : \forall i ∈ A \) and our algorithm 5 operates as unmodified AAT.

To this end, Figure 5.1 presents the accuracy reached by AAT with randomised preferences among peers, which is evaluated on scale-free networks with the expected degree \( ⟨d⟩ = 100 \) and other parameters as described earlier in Section 3.3. When \( x = 1 \) the accuracy of consensus, defined in Equation 3.6, is the lowest, since the underlying communication network becomes disconnected and many agents never receive an opinion from which to form their own. For \( x = 5 \ldots 50 \) the accuracy closely follows the accuracy of a single sensing agent, \( R_{min2} \), indicating that the distributed opinion aggregation is absent. The smaller values of \( x = 3 \ldots 10 \) often result in disconnected topologies, where agents have to rely on a single opinion source. However, the accuracy of consensus for larger values of \( x = 10 \ldots 50 \) is more affected by a different drawback of the randomised preferences strategy. Specifically, the agents randomly select a new set of preferred peers.
Chapter 5 Accurate Consensus with Identified Peers

![Graph](image.png)

**Figure 5.1:** The accuracy of consensus depending on the connectivity threshold with randomised preferences among peers. Here and in the all following results the shaded area represents the standard error of the mean. Additionally shown are the maximum of the accuracy, $R_{\text{max}}$, and the accuracy of a single sensing agent, $R_{\text{min2}}$.

after each opinion sharing round, and thus the network topology dramatically changes. In these circumstances, the AAT algorithm has to assign high weights to peers in order to guarantee the opinion formation. With such high weights the system operates in the unstable mode of its collective behaviour, when the agents share early opinions on a large scale and thus the distributed aggregation is absent. Finally, for large thresholds, $x > 50$, the network topology becomes stable enough for AAT to establish the critical mode of behaviour in the system and significantly improve the accuracy. This value of the connectivity threshold, $x$, approaches the expected connectivity of the network in our experimental setup, $\langle d \rangle = 100$. Thus, we can conclude that randomised preferences among peers results in a negative impact on the accuracy.

Considering the result, we design another benchmark indicator of peer importance, which requires external knowledge of the network topology.

5.2.2 Preferences by the Shortest Path Length to Opinion Sources

Earlier in Section 2.2 we discussed the principal influence of network topology on the dynamic processes in societies and, in particular, analysed this in the context of our model in Section 3.2. The main structural property of a single agent $i$ is the shortest path length to other agents, $l_{ij}$, in particular to the sensing agents, $j \in S$. The problem of finding these shortest paths in a decentralised fashion is well studied in the field of adaptive traffic routing in large decentralised networks, such as the Internet (Wolpert et al., 1999) or transport networks (Arokhlo et al., 2011).

However, in order to calculate the shortest path length, the agents have to communicate additional service messages, or rely on knowledge of the network topology. Both of these options are prohibited by our research requirements, and thus excluded in our model definition, by restricting the local view of an agent and limiting communication.
to opinion sharing only. Considering these requirements, we offer the indicator of peer importance based on the shortest path length only as a benchmark. It enables us to investigate how changes to network topology affect accuracy and subsequently to look for other indicators that may lead to the same result.

More specifically, we offer two indicators of peer importance based on the shortest path length to the opinion sources:

- the highest preferences to the nearest peer to *any* sensing agent:
  \[
  \phi_{ij}^{p\text{-any}} = -\min \{l_{ij,s} : s \in S\} 
  \]
  (5.1)

- the highest preferences to the nearest peer to *all* sensing agents:
  \[
  \phi_{ij}^{p\text{-all}} = -\frac{\sum_{s \in S} l_{ij,s}}{|S|} 
  \]
  (5.2)

Intuitively, by connecting to the nearest opinion sources, the agents reduce the risk of the double counting fallacy by aggregating only the earliest opinions. At the same time, by connecting to a few sensing agents, they are expected to reach a higher accuracy than that of a single sensing agent, $R_{\text{min}2}$. Note, that we cannot consider cases when agents select the longest distance to the opinion sources, since such strategies result in a disconnected network without multi-agent coordination.

We evaluate these indicators by replacing the definition of the preferences on the first line of Algorithm 5. On a small scale both of these metrics result in a similar topology illustrated in Figure 5.2. Here we experiment with a grid topology limiting the connectivity threshold to $x = 2$. As the result, the agents form a new topology from the existing one, which has sensing agents as hubs in its centre. The new topology has well connected groups of agents and thus can benefit from the presence of several sources of
new opinions, thereby forming a more accurate consensus than the accuracy of a single sensing agent, $R_{min2}$.

Figure 5.3: The accuracy of consensus depending on the connectivity threshold with preferences among peers as the shortest path length to the sensing agents ($\phi_{ij}^{p\text{-}any}$ and $\phi_{ij}^{p\text{-}all}$). Additionally shown are performance for the randomised preferences, $\phi_{ij}^{\text{rand}}$; the maximum of the accuracy, $R_{max}$, and the accuracy of a single sensing agent, $R_{min2}$.

Unmodified AAT delivers the same level of accuracy as $x > 140$.

Figure 5.3 presents the evaluation results for our experimental setup. Given the same connectivity threshold $x$, the indicator based on the shortest path length to any sensing agent, $\phi_{ij}^{p\text{-}any}$, outperforms the average shortest path $\phi_{ij}^{p\text{-}all}$. Crucially, it shows that the same level of accuracy can be achieved by limiting the connectivity from $x = \langle d \rangle = 100$ to $x = 10$.

However, for the small connectivity threshold, $x < 10$, the shortest path indicators exhibit lower performance than the randomised preferences indicator. Analysis shows that in these circumstances both indicators, and especially $\phi_{ij}^{p\text{-}all}$, occasionally lead to disconnected groups of agents forming out-of-loop connections in large networks.

Since AAT with the shortest path indicator, $\phi_{ij}^{p\text{-}any}$, approaches its maximum level of accuracy with a significantly limited connectivity, we can conclude that it is a promising indicator of peer importance. However, as we noted earlier, it cannot be used under our research requirements and thus it only acts as a benchmark for the indicators that are based on the local view of an agent.

### 5.2.3 Preferences by Opinion Timeliness

The first indicator of peer importance which meets our research requirements is the timeliness of a received opinion. By analysing the local view of agent $i$ described in Section 5.1.3, we note that all variables are dynamic in time, which is represented as timestep $k$ in terms of our model. Intuitively, the time when a peer communicates its
opinion indicates its distance to the sensing agents. We can consider two cases of opinion timeliness:

1. If the peer is expected to communicate its opinion earlier than others, then it is closer to the sensing agents. Thus, such an indicator implements a myopic approximation of the shortest distance to any sensor, as we discussed above. Formally we define the *earliest opinion indicator* as the following:

\[
\phi_{ij}^{t-\text{early}} = -E \left[ \min \left( \{ k : o_k^j \} \cup \{ k_{\max} \} \right) \right]
\]  

(5.3)

where \(E[\ldots]\) is the expected value of the earliest timestep when peer \(j\) communicates its opinion. The \(k_{\max}\) is largest timestep in the opinion sharing round, which is included in order to penalise peers which do not report their opinions. This expected value is learned over a number of opinion sharing rounds, and makes it more preferable to connect to peers that are the earliest to communicate their opinions.

2. Conversely, if the peer is expected to communicate its opinion is the last one, it is the most distant to the sensing agents. We define the *latest opinion indicator* simply as the opposite to the previous indicator with a default value of 0 to penalise peers which never communicate their opinions:

\[
\phi_{ij}^{t-\text{late}} = E \left[ \max \left( \{ k : o_k^j \} \cup \{ 0 \} \right) \right]
\]  

(5.4)

We did not directly implement such an indicator in the previous section because if all agents only prefer peers with the longest path length to the sensing agents, then they form a disconnected network. Since a multi-agent coordination to solve this problem cannot be considered due to the restricted communication in our settings, we skipped its analysis. In this case, with the latest opinion indicator as a substitute for such metrics, the agents self-organise into long sharing paths. At the same time, they guarantee their connectivity to the sensing agents given the additional clause in the indicator. Intuitively, when an agent informs its beliefs with the latest opinions that arrive to its neighbourhood, it ensures that these opinions are the result of a number of aggregations of other agents. Similarly, just as our society’s reviews of past events are often aggregates from a number of news sources, the agent may expect that the latest opinion is more accurate than the earliest opinion received directly from a sensing agent.

Now we test our hypotheses by evaluating these indicators on opinion timeliness in comparison to the benchmarks we introduced earlier. The topology produced by disconnecting the least preferable agents according to these indicators are illustrated in Figure 5.4. The topology produced by the latest opinion indicator forms clear groups of agents. However, the latest opinion indicator still encourages direct connections to the sensing
agents, thus connecting all these groups to the opinion sources. This confirms our hypothesis that it self-organises the system into long sharing paths whilst still ensuring its connectivity with the sensing agents.

Moreover, the evaluation of both timeliness indicators with regards to our benchmarks is presented in Figure 5.5. As it can be seen, both indicators deliver a higher level of accuracy than the randomised preferences benchmark and perform close to the indicator based on path length to the sensing agents. Crucially, both indicators deliver significantly higher performance for the system with the extremely low connectivity of $x = 1 \ldots 3$. This indicates that, unlike in our benchmarks, the self-organised topologies formed by these indicators are connected to the sensing agents. The earliest opinion indicator performs closely to the shortest path indicator, confirming our intuition that it is
a good approximation of the shortest distance to any sensor within the local view of an agent. Finally, the latest opinion indicator significantly outperforms the earliest opinion indicator and the benchmarks. This result confirms our reasoning in designing this indicator; that the latest opinions are more accurate than the early ones. This makes the latest opinion indicator a promising solution for determining preferences among agents’ peers.

However, our model does not define when an opinion sharing round finishes, by which the latest received opinion could then be clearly identified. So we introduced the timestep, \( k \), in order to simplify the modelling process. Specifically, we relied on the assumption that all agents can synchronously start a new opinion sharing round by resetting their opinions to the initial undetermined state. In order to avoid building a solution to our research problem which relies on assumptions that may not hold in realistic scenarios, we now investigate another indicator of peer importance.

### 5.2.4 Preferences by Opinion Surprise

In this section we develop an indicator of peer importance which, instead of analysing the time when peers communicate their opinions, focuses on their information content. The field of information theory (Shannon, 1948) offers us tools to analyse such information dynamics and content without restriction to a specific domain. Specifically, discrete opinions and the probabilistic private beliefs of the agents are means to quantify information, and thus are subjects of information theory. Crucially, these concepts are introduced directly from our motivating scenarios and are general to all potential applications. Thus, unlike in the previous section, in designing an indicator based on such an approach we do not need to rely on the assumptions we made in our model design.

In more detail, information theory introduces several key measures of information. The first one we focus on is the self-information which measures the uncertainty associated with a single interaction outcome. It is usually expressed by the average number of bits needed for storage or communication of the information. One bit of information is enough to answer a question without prior beliefs, such as “which opinion is correct, orange or blue”? In our model of an agent the “answers” are the opinions which the agent receives from its peers. The possible opinions \( o_k^j \in \{\text{orange}, \text{blue}\} \) have the corresponding probabilities of being correct in terms of the agent’s private belief: \( \{p_k^j, 1 - p_k^j\} \). Specifically, information theory quantifies the self-information \( I \) associated
with each received opinion $o_j^k$ as follows:

\[ I(o_j^k) = -\log_2 p_i(o_j^k), \]  

(5.5)

\[
p_i(o_j^k) = \begin{cases} 
  p_i^k & \text{if } o_j^k = \text{orange} \\
  1 - p_i^k & \text{if } o_j^k = \text{blue}
\end{cases}
\]

where $p_i(o_j^k)$ is the private belief of agent $i$ that the received opinion, $o_j^k$, is correct (following the definition that the agent’s belief, $p_i^k$, is the probability that the correct opinion is orange, see Equation 3.2). As the private belief, $p_i^k$, is always less than one, so $I(o_j^k)$ is always positive. When the agent believes that orange is the wrong opinion, $p_i^k$ becomes smaller, so the information content brought with such an opinion $I(o_j^k = \text{orange})$ becomes larger. Essentially, if the received opinion contradicts the agent’s opinion (and thus its private belief), then such an opinion is much more surprising. Therefore, this measure has also been called surprisal (Tribus, 1961), as it represents the “surprise” of seeing the outcome, or a new opinion in our model.

However, the surprisal of a single opinion does not solve our problem of determining preferences among peers. For this, we turn to the next key measure in information theory. The Shannon entropy rate of peer $j$, denoted as $H_{ij}$, quantifies the expected value of the information contained in each opinion $o_j^k$ received from peer $j$:

\[ H_{ij} = E\left[I(o_j^k)\right] \]  

(5.6)

If every new opinion received from peer $j$ contradicts the belief of agent $i$, and thus, is very surprising, the entropy rate $H_{ij}$ holds its maximum value 1. Conversely, if the peer always communicates an opinion which follows the private belief of the agent, $H_{ij}$ approaches its minimum value 0.

Given this, we adopt the entropy rate as the indicator of peer importance. In order to align our notation we denote it as follows:

\[ \phi_{ij}^{\text{surp}} = H_{ij} \]  

(5.7)

This opinion surprise indicator shows how surprising the opinions received from a peer are. It is minimised for peers which form their opinions following the agent’s opinion, since their opinions lack new information. The agent determining preferences by the entropy rate ignores such peers, and thus mitigates the double counting problem. In contrast, the agent assigns a higher preference to any sensing agent in its neighborhood, which originates new opinions and has a high entropy rate. Similarly, the agent can identify peers which form their opinions from a different sensing agent, and thus, significantly contribute to forming an accurate opinion.
Following our evaluation procedure, we now analyse the performance of the opinion surprise indicator. In particular, Figure 5.6 illustrates a sample network topology formed by this indicator. As it can be seen, all agents prefer to connect to the peers which are the closest to the sensing agents with an exception of a single agent, which does the opposite. It did not converge to the optimal solution before the simulation was interrupted, however this would not affect its performance considering that its peers are well connected to the same sources. Considering this, we can conclude that the indicator self-organises the system into a beneficial network pattern, which may improve the accuracy of consensus in dense networks of our full experimental setup.

The results presented in Figure 5.7 indicate that on the scale of systems we evaluate in our experimental setup, the formation of such network patterns leads to a high accuracy of consensus. The opinion surprise indicator outperforms the randomised preferences.
benchmark, and follows the results of the shortest path indicator for high values of the connectivity threshold. However, as the connectivity threshold becomes smaller, \( x = 4 \ldots 10 \), the new indicator clearly outperforms the benchmarks. Crucially, for \( x = 10 \) it exhibits the level of accuracy which is higher than that for any other connectivity threshold. This supports our earlier hypothesis that the informed limitation of the connectivity may mitigate negative effects of the double counting problem and result in a higher accuracy of consensus.

Considering the discussion of the indicators of the most important peers up to this moment, we conclude that the opinion surprise indicator is the most promising step towards determining preferences among peers. Thus, in next section we address the second stage of designing a behavioural algorithms which attributes individual weights to the agent’s peers.

5.3 The Individual Weight Tuning Algorithm

In this section we design the Individual Weight Tuning (IWT) algorithm, which is an adaptive behavioural algorithm for improving the accuracy of consensus. The adaptivity comes from the fact that unlike algorithms limiting the connectivity of an agent, which we discussed above, IWT does not require the additional connectivity parameter. Instead, IWT is the first algorithm which assigns weights individually to each peer of an agent. This is also its crucial difference to the AAT algorithm presented in the previous chapter.

Following our analysis of the indicators of the most important peers, IWT employs the opinion surprise indicator. Unlike the connectivity limitation benchmark, which applies the same weight towards selected peers, IWT distributes weights across all the peers of an agent. Specifically IWT attributes the highest weights to the most surprising peers, which are most likely to be communicating new opinions from the sensing agents. To the rest of the peers, which participate in the opinion sharing cascade and report the already aggregated opinions, IWT attributes low weights. Since there is no fixed threshold on the number of important peers, each agent running IWT individually decides how many of its peers report valuable opinions. This adaptivity of IWT enables agents to efficiently aggregate opinions regardless of the number of peers in their neighbourhood. Highly connected IWT agents aggregate opinions only from the few peers that are nearest to the sensing agents, thus forming a more accurate opinion and communicating it to the rest of the peers in their neighbourhood. Whilst IWT agents connected to a small number of peers are likely to focus on the single peer originating opinion cascades in their neighbourhood. Both extremes contribute to the accuracy of consensus, through better-informed decision making by highly connected agents whose opinions are then shared in cascades, without generating overconfident beliefs.
On the system level, in order to improve the accuracy of consensus, IWT employs the same technique as the AAT algorithm by tuning the system into the critical mode of collective behaviour. More specifically, IWT incorporates AAT as a part of its design. In doing so, IWT reaches the critical mode relying on the same, already-evaluated, indicator: the awareness rate of an agent.

However, the search space of IWT is different. Instead of the candidate weights $W_i$ of agent $i$ in AAT, IWT generates set $S_i$ of susceptibility levels, which are scaling factors for the entropy rates of the peers. The rest of the AAT procedure is preserved as described in Section 4.1. Over a number of opinion sharing rounds, the algorithm discovers the peers’ entropy rates and the susceptibility level to be used. The latter factor encodes how responsive the agent should be to incoming opinions, similar to susceptibility in the modelling of infectious diseases. More specifically, the weights $w_{ij}$ agent $i$ individually attributes to its peers $j \in D_i$ are calculated by IWT as follows:

$$w_{ij} = \frac{\phi_{ij}^{\text{surp}}}{\max(\{\phi_{il}^{\text{surp}} : l \in D_i\})} \cdot s_i + 0.5\quad (5.8)$$

where $\phi_{ij}^{\text{surp}}$ is the entropy rate of peer $j$ calculated as described above in Section 5.2.4 and divided by its maximum value in order to scale the range of preferences among peers to $[0, 1]$. The second multiplier, $s_i$, is the currently used susceptibility level and is calculated in the same way as the weight out of the candidates by regular AAT. The search space of the candidate susceptibility levels, $S_i$, is populated with values drawn from the range $[0, 0.5]$ with a given step size, for example 0.01. The last summand of 0.5 ensures that the resulting weight $w_{ij}$ is in the range of $[0.5, 1]$.

Figure 5.8: Sample weights selected by the IWT algorithm. For the first time peers of each agent are connected by individually selected weights.

Figure 5.8 illustrates how weights are defined by IWT in a small system. Unlike in all previous algorithms, the weights attributed by the agent to each of its peers are different. This is represented by the variable width of the connections. All agents
clearly prefer peers that are closer to the sensing agents or stand in a sharing path to a more distant sensing agent. Such a connectivity pattern on the system level follows our intuition, as discussed, alongside the opinion surprise indicator, which is used to determine preferences among peers.

To summarise, the activity diagram of IWT is shown in Figure 5.9. Similar to AAT, before operating IWT initialises its search space (step 1), which is the set of susceptibility levels. Then, after each opinion sharing round, it calculates the awareness rates for all susceptibility levels given the old entropy rates of the peers (step 3a) and selects a new level to use (3b). In parallel, during the opinion sharing round, or after, IWT calculates the entropy rate of each peer, which determines the preferences among peers (step 2). Finally, IWT updates the weights it attributes to the agents’ neighbours following Equation 5.8 (step 4). After a number of opinion sharing rounds IWT converges to a stable set of weights.

IWT is based on the AAT algorithm and it has the same single parameter, the target awareness rate $h_{\text{trg}}$. Since we only changed the search space of the algorithm, the
optimisation problem of selecting the current weight by AAT, or the susceptibility level in the IWT design, remained unmodified. Therefore, the results of the analysis of AAT parameters we conducted in Section 4.3 can be extended to IWT. Specifically, we base the IWT search procedure on the hill climbing strategy and adopt the same target awareness rate $h_{trg} = 0.9$. Having defined the parameters of the algorithms as such, we can conduct its empirical study.

5.4 Empirical Evaluation

In this section we empirically evaluate the IWT algorithm in order to examine its compliance with the research requirements. We benchmark its performance against our AAT algorithm, since, as we showed in the previous chapter, it outperforms the existing solution, DACOR. Additionally, we compare the performance with a number of pre-tuned benchmarks that were introduced in Section 3.5. The experimental setup is carried over from the model and AAT evaluations.

In the following subsections we analyse our performance metrics: the accuracy of consensus achieved by the IWT algorithm in Section 4.4.1; its communication expenses in Section 4.4.2; and its computational expenses in Section 4.4.3.

5.4.1 The Accuracy of Consensus

The experimental setup is summarised in the previous chapter in Table 4.1. However, unlike in the AAT analysis where we focused on sparse networks, here we focus on comparing performance between sparse and dense networks. This is due to the fact that AAT performed on a par with our benchmarks in sparse communication networks, and so a more challenging setup is required. Moreover, as we discussed earlier in Section 5.1, attributing individual weight to peers is much more beneficial when agents face an abundance of reported opinions. In contrast, in the sparse topologies with degree $\langle d \rangle = 8$ that we used to evaluate the AAT algorithm, agents tend to form their opinions after receiving on average less than 2 opinions from their neighbours. Therefore, considering it is this deficit of peers that actually triggers an opinion change, differentiation of the peers is not expected to bring additional value. Thus, IWT is expected to exhibit similar performance to AAT on sparse networks.

In Figure 5.10 in the left column (Figures A, C and E) we compare the accuracy of consensus in systems running AAT and IWT behavioural algorithms on sparse topologies with the pre-tuned benchmarks (see Section 3.5). The results showing that IWT delivers similar level of the accuracy as of AAT. This finding supports our hypothesis that determining preferences among peers in sparse networks does not bring a competitive advantage.
Chapter 5 Accurate Consensus with Identified Peers

Figure 5.10: The accuracy of consensus achieved by the IWT algorithm in comparison to the benchmarks and the AAT algorithm. Network size, $N$, network topology and the expected degree, $\langle d \rangle$, are variables in this setup. (All agents are running the Bayes aggregation function, see Appendix C for the Weighted sum aggregation function)
It is important to notice, however, that IWT delivers close results to AAT. This confirms that IWT shares the properties of the AAT algorithm it is built on. Moreover, IWT outperforms AAT in the case of scale-free topology, which contains few agents with a large number of peers, so-called “hubs”. Through applying individual weights towards peers, IWT enables hubs to form more accurate opinions. Since these hubs are critical in sharing processes in scale-free networks, the accuracy of consensus notably increases.

In Figure 5.10 in the right column (Figures B, D and F) we present results for dense communication networks \(\langle d \rangle = 100\). Crucially, the IWT algorithm outperforms AAT on all network topologies and sizes. This result confirms that IWT is superior in these settings, when agents face a large number of opinions from their peers and have to determine preferences among them. In delivering such results, which are very close to our centrally pre-tuned benchmarks, IWT outperforms the state-of-the-art algorithms AAT and DACOR, and provides the most promising solution to our research problem.

Comparing the performances of IWT and AAT, we can conclude that IWT is significantly more beneficial in systems with dense communication networks whilst AAT performs well on sparse networks. Additionally, we can conclude that IWT is adaptive to changes in the experimental settings. In order to investigate scalability we now investigate its communication and computational expenses.

### 5.4.2 Communication Expense

The IWT algorithm is designed not to introduce additional communication above that already present. Specifically, our research requirements impose strict limitations on communication by preventing agents from sharing any additional information besides their opinions. By analysing the number of shared opinions in the system controlled by IWT, we can make conclusions about its scalability. As a base line we defined minimal communication, \(U_{\text{min}}\) (Equation 3.12), as the number of messages required to share an opinion between all agents in the system in a single opinion cascade.

Figure 5.11 presents communication expenses as the number of messages required to share all opinions in the system during an opinion sharing round. The results for IWT and AAT follow each other very closely and overlap on the scale of the plot. This confirms that IWT tunes the system to the same state of opinion dynamics and inherent properties as AAT.

The communication expenses of IWT are slightly lower than the minimal communication, \(U_{\text{min}}\). This can be explained by the fact that during some of the opinion sharing rounds the agents do not reach a consensus and opinions are not disseminated on a large scale, thus decreasing the average value of the number of communicated messages. Crucially, the fact that IWT expenses closely follow minimal communication confirms the scalability of IWT.
Chapter 5 Accurate Consensus with Identified Peers

5.4.3 Computational Expense

Finally, to investigate the efficiency of IWT, we measure its computational cost. Since IWT inherits the properties of AAT and changes agents’ weights only once at the end of each sharing round, we omit analysis of the weight changes as we did for the DACOR algorithm. In this evaluation we focus on measuring the runtime of IWT in comparison to: i) a system running without a behavioural algorithm (with individually pre-tuned fixed weights); ii) AAT; and iii) its computationally more efficient version, iAAT.

Figure 5.12 presents results indicating that IWT is significantly more resource consuming than AAT across all setups, and an order of magnitude slower than iAAT or a simulation of a system with fixed weights. These results confirm our concerns that IWT has to analyse the significantly larger search space, which results in the performance penalty.

However, we have to notice that most of the computations are done during the simulation of the opinion sharing round. In contrast, AAT runs only once at the end of the round. Specifically, IWT running agents update the entropies of their peers as they receive opinions from them. Thus the indicated increase in computational complexity does not
require high performance agents, since the computations are evenly spread throughout the opinion sharing round. This enables us to conclude that despite the significantly higher computational cost, IWT still suits our research scenario and can operate on agents with limited resources.

Finally, the memory cost of IWT is higher than that of AAT, since it has to calculate individual weights to each neighbour. Specifically, IWT requires us to calculate and store the entropy rate of each neighbour, $H_{ij} \forall j \in D_i$, and a weight attributed to it, $w_{ij} \forall j \in D_i$. Additionally, IWT stores the set of candidate susceptibility levels, $S_i$, with the corresponding awareness rates delivered by each candidate, $\hat{h}(w_i^l) \forall s_i^l \in S_i$. In our experimental setup above, the average number of network neighbours $\langle d \rangle = 100$ and the set of candidates is populated with $|S_i| = 50$ values. Thus, the average memory consumption per agent is $\langle d \rangle \cdot 3 + |S_i| \cdot 2 = 400$ values. However, we have to consider that in scale-free networks some of the agents, so called network hubs, have magnitudes higher number of neighbours than average. Thus, in large scale-free networks of our experimental setup several highly connected agents may require up to $4000 \ldots 10000$ values to store.

Figure 5.12: Computational expenses for the IWT algorithm as time required to simulate a system. Here we provide comparison against the AAT and iAAT algorithms which differentiate in their computational expenses, and against an simulation of a system with fixed weights.
Chapter 5 Accurate Consensus with Identified Peers

5.5 Summary

In this chapter we presented the Individual Weights Tuning (IWT) algorithm, which is the first solution to differentiate between network neighbours of an agents, in solving the problem of improving the accuracy of consensus in a decentralised fashion. Specifically, IWT is the first behavioural algorithm that adjusts weights for each opinion source individually. In contrast, the AAT algorithm, presented in the previous chapter, and DACOR, discussed in Section 2.4, assign the same weight to all the peers of an agent.

Crucially, due to this unique feature, the IWT algorithm outperforms the existing decentralised solutions for systems with dense communication networks. In these circumstances, determining preferences among agent’s peers brings clear benefits. Agents running IWT assign higher weights to the peers which deliver the most surprising opinions. Such peers are closer to the sensing agents in the opinion sharing path. Therefore, by ignoring opinions reported by the rest of the peers, which are further along this sharing path, agents are less likely to double count the same opinions. This implements a more accurate opinion aggregation and results in a higher accuracy of consensus. However, in sparse networks the number of opinions received by an agent is small, and thus, determining preferences among peers does not introduce noticeable benefit, in which case AAT is a better solution.

We empirically evaluated the IWT algorithm in order to investigate its compliance with the research requirements. Additionally we benchmarked it against our own algorithm, AAT, as presented in the previous chapter. As a result, we showed that IWT meets the research requirements identified in Section 1.4 by delivering high levels of:

1. **Accuracy**: IWT significantly outperforms AAT in all experimental setups with dense networks. At the same time, the accuracy of consensus delivered by IWT is very close to that of the AAT algorithm in sparse networks. This is explained by the fact that in sparse networks agents receive on average less than 2 opinions before forming their own opinion, and thus, determining preferences among the opinion sources does not introduce much benefit. Therefore, it is clearly more beneficial to apply IWT to systems with dense networks, whilst AAT is the preferable solution for sparse networks due to its significantly higher computational efficiency.

2. **Adaptivity**: Similarly to AAT, IWT exhibits high adaptivity by delivering a high accuracy of consensus when compared with the benchmarks in all our experimental setups with variable network topologies and densities.

3. **Scalability**: Again similarly to AAT, IWT scales well with system size. This is due to having the same approach to improve the accuracy of consensus, by exploiting the properties of collective behaviour. Specifically, systems of 5000 and 10000
agents achieve close levels of accuracy showing its maximum performance. Moreover, since communication and computational expenses per agent stay constant, we conclude that the solution is highly scalable.

4. **Communication efficiency**: This requirement is met by the IWT design and does not introduce additional communication above what is already present. Moreover, since IWT determines preferences among peers, it creates opportunities to reduce communication even further. Specifically, if a system operates on a directed dynamic network and agents are able to disconnect from their peers, IWT can significantly reduce communication by disconnecting subscriptions to the least important peers.

5. **Computational efficiency**: The resources to run IWT are significantly higher than that required for AAT. However, unlike AAT which runs only once at the end of each opinion sharing round, most of the IWT calculations are performed during the sharing round. Specifically, IWT dynamically updates estimates on peers’ entropy rates with every received opinion, which is responsible for the increase in its computational expenses. Crucially, despite the significantly higher number of weights that IWT is required to tune on the system scale in order to reach the desired state in the collective behaviour, it still converges to the solution in the same number of sharing rounds as AAT.

Considering the contributions listed above, we extend the state-of-the-art in solving the problem of reaching accurate consensus in large systems with restricted communication. Our evaluation showed that the IWT algorithm is particularly beneficial when applied to systems with dense networks, where determining preferences among opinion sources is more critical. Together with the AAT algorithm, which is more computationally efficient and therefore more suitable for sparse networks, we meet all our research requirements.
Chapter 6

Conclusions

In this chapter, we review the contribution of this thesis towards our research aim of achieving accurate consensus in large multi-agent systems with restricted communication. In particular, in Section 6.1 we summarise the research carried out and explain how each contribution has satisfied the design requirements laid down in Chapter 1. Then, in Section 6.2 we identify several potential lines of future research that could be pursued as a continuation of this work.

6.1 Conclusions

In this thesis we argued that achieving accurate consensus in large multi-agent systems is an important problem. Specifically, we posited the need for solving this problem in the challenging setting of restricted communication, in which agents are only able to share their opinions without any supporting information. In order to approach this problem, we identified three research challenges that need to be addressed:

1. how to influence the accuracy of the consensus that relying on the collective behaviour in large multi-agent systems with restricted communication;
2. how to induce the desired mode of collective behaviour in a decentralised fashion in anonymous networks, where an agent cannot differentiate between its peers;
3. how to improve the accuracy even further in networks where agents can identify the sources of each opinion they observe.

In tackling these research challenges, our work has advanced the state-of-the-art in the field of emergent behaviour in multi-agent systems. Specifically, we contributed to opinion formation modelling and, crucially, developed novel methods for improving the accuracy of consensus by exploiting the properties of collective behaviour. In the following
paragraphs, we summarise each chapter in relation to how it tackles the corresponding research challenge.

After reviewing the existing research on opinion sharing in large systems in Chapter 1, and in further detail in Chapter 2, the properties of collective behaviour were chosen as the point of departure for this work. This is because the traditional solutions for achieving accurate consensus require additional communication in order to operate, which violates the communication restriction in our research problem. For example, the agreement protocols introduce a large number of interactions and any algorithm based on reasoning about the accuracy of communicated information requires additional annotation. Therefore, we focused on investigating how different dynamics of opinion sharing in large systems affect the accuracy of consensus. Specifically, we chose agent-based modelling in order to study realistic settings, such as the different decision making processes employed by the agents, and complex communication networks, which are known to have a significant effect on dynamic processes. However, none of the existing models addressed all these aspects in the context of our research challenge. Moreover, to date, none of the reviewed research has solved our problem directly.

Against this background, building on the existing research in Chapter 3 we designed a new opinion sharing model which addresses the existing shortcomings. Our model is a large system of agents connected by a network with complex topology, with only a few sensing agents which make noisy observations, and thus, dynamically introduce conflicting opinions into the system. Each agent informs its private beliefs by observing the opinions of its network neighbours, or peers, and after forming its own opinion re-shares it with them. Since the correct opinion slightly predominates in the observations, the system is likely to converge to the correct consensus, and the expected probability of such an event defines the accuracy of the consensus.

Crucially, our model is the first to quantify the impact of collective behaviour on the accuracy of the consensus. In the analysis of our model, we extended the existing findings on linking collective behaviour to opinion sharing. Specifically, we investigated the properties of our model and showed the existence of a narrow range of the critical parameters in which incorrect opinions are filtered out during the sharing process. With these critical parameters the system achieves settings of distributed opinion aggregation, and thus, benefits from the presence of a number of sensing agents by exploiting the properties of the system’s dynamics. Specifically, early, and possibly inaccurate opinions, are shared in cascades amongst small groups of neighbouring agents to prevent overreaction. Only when several groups with the same opinion overlap is this locally-supported opinion disseminated in a large cascade thereby leading to consensus. Such collective behaviour results in a significant accuracy improvement in comparison to the accuracy of a single sensing agent. However, we showed that due to a different objective set up in our research, the critical parameters which induce such collective behaviour do not coincide with the predictions made for the existing model. Therefore, we addressed this gap
and analysed the properties of this critical mode of collective behaviour. As the result of this, we suggested which properties of system dynamics indicate whether the system is operating in critical mode, and crucially, are invariant to the system parameters such as its size and the topology of the communication network. This contribution provided an answer to the first research challenge.

To take a step further and improve the accuracy of consensus by exploiting these properties of collective behaviour, in Chapter 4 we develop the Autonomous Adaptive Tuning (AAT) algorithm. This is the first algorithm which tunes the system into the critical mode of collective behaviour in settings where communication is strictly limited to opinion sharing. Relying only on the observation of local opinion dynamics, each agent running AAT gradually regulates a weight it attributes to its peers. This weight represents the influence of the opinions it receives from its neighbours on its own opinion, thus encoding the local opinion dynamics. As a result of applying AAT, 80-90% of the agents in a large system form the correct opinion, in contrast to 60-75% for the state-of-the-art message-passing algorithm, DACOR, proposed for this setting. Also, we confirmed that AAT is the first solution that operates with the minimal communication requirement and is computationally inexpensive, while DACOR requires a significant communication overhead and considerably higher computational cost. Additionally, we test other research requirements and demonstrate that AAT is both scalable and adaptive by evaluating teams with different sizes and network topologies. Finally, we showed that AAT is highly robust since it significantly improves the accuracy of consensus even when only being deployed in 10% of the agents in a large heterogeneous system. Thus, with the proposed algorithm we have been able to solve the second research challenge.

In Chapter 5 we tackled the last research challenge and discussed how the agents can benefit from identifying the sources of received opinions, namely their peers. In Chapter 1 we argued that in modern information systems agents are often face an information overload and its filtering is essential in order to form an accurate opinion. In terms of our model, we analysed how preferences should be determined among peers, and showed that the expected entropy of a peer’s opinion, or its opinion surprise, is a promising indicator of its importance. Following this, we presented the Individual Weights Tuning (IWT) algorithm, which differentiates between the peers of an agent in solving the problem of improving the accuracy of consensus. Specifically, IWT is the first behavioural algorithm that adjusts weights for each opinion source individually. Agents running IWT assign higher weights to the peers which deliver the most surprising opinions. Such peers are closer to the sensing agents in the opinion sharing path. Therefore, by ignoring opinions reported by the rest of the peers, which are further along this sharing path, agents are less likely to double count the same opinions. This implements a more accurate opinion aggregation than applying the same weight to all peers, as in the case with AAT, and results in a higher accuracy of consensus. Crucially, by incorporating the information
about the source of an opinion, IWT outperforms AAT for systems with dense communication networks. Whilst in sparse networks IWT exhibits similar performance to AAT, since the number of opinions received by an agent is small and thus, determining preferences among peers does not introduce noticeable benefit. Considering that IWT has higher computational cost than AAT, due to its larger search space of the optimised weights, we conclude that IWT is more beneficial to use in dense networks, while AAT delivers a similar level of accuracy improvement in sparse networks but with a lower computational cost.

More specifically, looking back at the research requirements we identified at the beginning of this thesis, we can conclude that we have successfully addressed each of them:

1. **Accuracy**: For the AAT and IWT algorithms we gave extensive empirical evidence on the achieved accuracy of consensus by comparing them against: the theoretical bounds on the accuracy; the state-of-the-art algorithm DACOR; and the static benchmarks; based on a resource intensive empirical exploration of system performance with different parameters. Crucially, the algorithms outperform the existing solution, significantly improve the accuracy of consensus in comparison to the accuracy of a single sensing agent for systems with more than $N \geq 200$ agents and, finally, achieve a level of accuracy comparable to systems pre-tuned for the highest accuracy.

2. **Adaptivity**: We confirmed high adaptivity of our algorithms by evaluating them in a wide range of experimental setups by varying the system size, the network topology and its density, and the aggregation function used by each agent. Crucially, AAT and IWT do not require additional tuning for a specific domain and have a single parameter which is fixed across all experiments.

3. **Scalability**: We evaluated systems with up to 10000 agents, being limited only by the high computational expenses of the pre-tuned benchmarks. The AAT and IWT algorithms proved their scalability and showed that the level of accuracy improvement rises with the size of the system. This comes from their design, which is based on exploiting the collective behaviour which is more distinct in large systems, however it is not noticeable in systems less than $N \leq 100$ agents. At the same time, the computational and communicational costs for each individual agent remain constant. As a result, we believe that AAT and IWT can be used in much larger systems.

4. **Minimal communication requirement or Communication efficiency**: This requirement is met in the design of our algorithms. In particular, neither of them introduce additional communication above that already present in the system. Moreover, we showed that the communication exchange in the system controlled by AAT or IWT is in a range of error from the minimally-required communication
to share a single opinion between all agents. Conversely, the DACOR algorithm has several magnitudes higher communication expense since it communicates additional service messages in order to operate.

5. **Computational efficiency**: Crucially, the computational cost of our algorithms does not depend on the scale of the system and scales linearly with the size of the system. In order to reduce the computational cost for each agent, we designed the improved AAT (iAAT) algorithm which limits its search space to the set of optimal candidates’ weights. Importantly, the computational cost of iAAT is close to the cost of running a system without any behavioural algorithm. In comparison, the cost for AAT is 2-8 times higher whilst, due to a significantly larger search space, IWT is computationally the most expensive and requires 4-10 times longer simulation time than iAAT. However, we showed how computations of IWT may be distributed over the course of the opinion sharing round, thus it could be executed by agents with limited resources. Crucially, despite the significantly higher number of weights that IWT is required to tune, it still converges to the solution in the same number of sharing rounds as AAT.

6. **Robustness**: Finally, we showed that AAT is highly robust and does not require all agents in the system to participate in the optimisation process. In order to significantly improve accuracy, AAT need only be deployed on a small random subset (10-25%) of the agents in a heterogeneous system.

To summarise, in the context of restricted communication in large multi-agent systems, we have linked the collective behaviour to improvement of the accuracy of consensus, and identified its properties, in Chapter 2. Then, relying on the discovered properties of collective behaviour, we contributed a solution for decentralised accuracy improvement in anonymous networks in Chapter 3. Finally, we extended this solution for a case in which agents can identify the sources of received opinions and contributed an algorithm which can benefit from that in Chapter 4. Against this background, we can justifiably claim that we have addressed all of the challenges in the space of our research problem.

### 6.2 Future Work

As the discussion above suggests, the research presented in this thesis constitutes a significant step towards improving the accuracy of consensus in real-world applications. However, despite these accomplishments, there are still a number of open issues to be addressed. Specifically, our contributions were made and evaluated through the model of a cooperative and static environment, where agents do not have their own preferences and thus, do not compete. Therefore, future work should focus on addressing these limitations.
More specifically, we identify three lines of investigation to extend the scope of the applicability of our work:

- **Attack resistance**
  In this work we focused on cooperative multi-agent systems, and only considered the case of a heterogeneous system in order to analyse the robustness of our approach. However, the large multi-agent systems, which were discussed in the motivating section of our research, are often exposed to attackers which may actively manipulate agents’ opinions. For example, in a situation of conflict some agents might have malicious intentions and can act arbitrarily in order to compromise the existing local opinion. Given this, the next important direction of our work is to develop an attack-resistant solution that will help agents to reduce the negative influence of a small number of deliberate attackers.

  In more detail, this problem could be solved by developing an algorithm for detecting attacking agents in order to mitigate their impact on the accuracy of consensus. Efficient strategies of attacking agents for opinion sharing models and their impact on consensus were recently analysed by Glinton et al. (2011). Relying on the results of this work, it is possible to design new attack-resistant algorithms based on two approaches. Firstly, further research needs to investigate the efficiency of detecting attackers based only on agents’ local views in order to minimise communication. The agents in this approach would be making decisions based on learning the individual dynamics of their neighbours; a direct extension of the IWT algorithm we developed in this thesis. Secondly, a potential solution to the problem is enabling agents to change their neighbourhood dynamically. In doing so, they may increase their changes upon discovering the attacking agents, in comparison to others, and disconnect from them. Finally, there is a need for extensive comparison of the results against the existing literature on reputation in multi-agents systems (such as traditional Regret framework (Sabater and Sierra, 2002), or decentralised gossip-based approaches (Bachrach et al., 2008)), which offer algorithms ways to identify attackers using additional service messages or centralised authorities.

- **Improving opinion sharing in models based on game theory**
  Recently, opinion formation has been approached from the perspective of game theory. This lens offers opinion sharing models based on rational, self-interested agents with their own goals, such as compliance with majority opinion and establishing social order according to the so-called “norms” of a society (Grizard et al., 2007), or having their own preferences encoded as pay-off matrices given the opinions of their peers (Di Mare and Latora, 2007). Various preferences produce different kinds of opinion dynamics, some of which, however, are very similar in behaviour to our traditional opinion sharing model with a binary subject of interest (Ding et al., 2010).
Crucially, it was shown that heterogeneity in such types of evolutionary games on graphs, which are opinion sharing models, have a significant impact on a model’s dynamics. There is some similarity in this aspect of our model with the model offered by Yang and Wang (2010), in which each agent is assigned a weight encoding its influence. In this model, it is found that there is an optimal value of a parameter distributing these weights, which leads to the highest cooperation level or the fastest consensus. However, this parameter is tuned in a centralised fashion and has to be selected for particular model parameters. Such settings closely resemble ours, suggesting that the offered algorithms, AAT and IWT, can be transferred to a new domain, assisting in finding the optimised settings of collective behaviour in game-theoretical opinion sharing models.

- **Dynamic environments**

The final direction for future work is to study the applicability of the developed approaches to more complex settings, such as existing social communities or sensor networks. Specifically, the model of the environment used in this thesis contains a number of assumptions that allowed us to simplify its representation. Given this, we propose to extend it in order to capture additional features of realistic settings while maintaining the mathematical simplicity which allows for detailed analysis. In particular, there is a need to introduce:

- **Richer communication model**: The assumption we used in our model, that the subject of common interest expires after a certain deadline and the agents have to synchronously reset their beliefs and opinions, needs to be relaxed. To achieve this, the model can be extended in two main ways: (i) by enabling agents to reconsider their private beliefs over time without external synchronisation stimulus (i.e. opinion “forgetting”); or (ii) by enabling a simultaneous sharing of opinions about different subjects of interest.

- **Dynamic system**: The size of any realistic system and the topology of its communication network may change in time. We have discussed that the algorithm developed in this thesis can be applied in the dynamic settings in Section 4.1.2, however additional research is needed to examine this. In particular, the existing research on the evolution of social networks can suggest appropriate models for network growth, such as that offered by Price (1965) and Barabási and Albert (1999).

- **Dynamic agent**: Finally, network dynamics may be initiated by individual agents rewiring their connections. We approached this extension by enabling agents running IWT to ignore opinions from selected peers. Thus, each agent is able to reduce the number of its own connections. However, in order to find new peers and evaluate them, the agent has to be equipped with algorithms for exploring the network structure.
Crucially, each of these extensions has a significant impact on the dynamics of opinion sharing and thus, on the exact parameters of the emergent behaviour when the accuracy of the team improves. As the result, this may give us the ability to evaluate the adaptivity and the robustness of our solutions even further.

By meeting these challenges, we would be able to extend the practical applicability of the solutions developed in this thesis.
Appendix A

Additional Results for the Model Evaluation

In the following Figure A.1 we provide additional results which show dependency between the common critical weight, $w_c$, and the accuracy of consensus in the critical mode. This view enables us to compare the weights that lead to the critical mode in different network topologies and densities.

Figure A.2 shows the effective branching factor in the critical mode depending on the accuracy of consensus. This additional result is provided to testing the hypothesis suggested by Glinton et al. (2009) that branching factor is equal to 1 in the critical mode. High precision in a case with a random network might be a promising result, however, branching factor is not indicative measure of critical state for complex topologies. This experiment also explains high sensitivity to the settings of the existing solution for finding critical weights in a distributed fashion, DACOR algorithm, that we discussed in Section 2.4.

All additional experiments are conducted in the same experimental setup as the model evaluation, specifically we analyse systems of $N = 1000$ agents with variable network topology and density represented as the expected degree, $\langle d \rangle$. All agents are using the Bayesian aggregation function and final results are averaged over 50 iterations for each topology instance.
Appendix A: Additional Results for the Model Evaluation

Figure A.1: Highest accuracy and agents weights in the critical mode. Horizontal error bars indicate on the range of $w_c$ which lead to at least 95% of the highest accuracy.

Figure A.2: Testing hypothesis that branching factor is equal to 1 in the critical mode of behaviour, when the highest accuracy of consensus is observed.
Appendix B

Additional Results for AAT Evaluation

In Figures B.1 and B.2 we provide additional results of the AAT evaluation. Specifically, we consider the experimental setup with the Bayesian aggregation function on a dense communication network in Figure B.1, and than based on the Weighted Sum aggregation function in Figure B.2. Despite our model exhibited different dynamics with the Bayesian and Weighted Sum aggregation functions (Section 3.3), AAT achieves the similar level of accuracy as with the Bayesian aggregation function.
Figure B.1: Accuracy of consensus achieved by AAT, DACOR and the benchmarks depending on the system size and topology ($\langle d \rangle = 100$, the Bayesian aggregation function)
Appendix B Additional Results for AAT Evaluation

Figure B.2: Accuracy of consensus achieved by AAT in comparison to the benchmarks and the DACOR algorithm. Network size, N, network topology and the expected degree, \langle d \rangle, are variables in this setup. All agents running the Weighted Sum aggregation function.
Appendix C

Additional Results of IWT Evaluation

In Figure C.1 we provide additional results of the IWT evaluation. Specifically, we consider the experimental setup based on the Weighted Sum aggregation function. Despite our model exhibited different dynamics with the Bayesian and Weighted Sum aggregation functions (Section 3.3), IWT achieves the similar level of accuracy as with the Bayesian aggregation function.
Figure C.1: Accuracy of consensus achieved by IWT in comparison to the benchmarks and AAT. Network size, $N$, network topology and the expected degree, $\langle d \rangle$, are variables in this setup. All agents running the Weighted Sum aggregation function.
References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


